



HAL
open science

Recherche biomédicale et journalisme en situation d'incertitude : validité des résultats de la recherche biomédicale et couverture médiatique

Estelle Dumas-Mallet

► To cite this version:

Estelle Dumas-Mallet. Recherche biomédicale et journalisme en situation d'incertitude : validité des résultats de la recherche biomédicale et couverture médiatique. Science politique. Université de Bordeaux, 2017. Français. NNT : 2017BORD0737 . tel-01682946

HAL Id: tel-01682946

<https://theses.hal.science/tel-01682946>

Submitted on 12 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE
Présentée pour obtenir le grade de
DOCTEUR DE L'UNIVERSITE DE BORDEAUX

Ecole Doctorale SP2 : Sociétés, Politique, Santé Publique
SCIENCES PO BORDEAUX

Discipline Science Politique

par **Estelle DUMAS-MALLET**

**RECHERCHE BIOMEDICALE ET JOURNALISME EN
SITUATION D'INCERTITUDE**

**Validité des résultats de la recherche biomédicale et couverture
médiatique**

Sous la direction de **Andy SMITH**
Et la co-direction de **François GONON**

Soutenue le 10 novembre 2017

Composition du jury

Mme Lucienne LETELLIER,

Directrice de recherche émérite, Université Paris Sud, *Présidente*

Mme Isabelle BOUTRON,

Professeur des Universités, Université Paris Descartes, *Rapporteur*

M. Tommaso VENTURINI,

Advanced Research Fellow, Ecole Normale Supérieure de Lyon, *Rapporteur*

M. Andy SMITH,

Directeur de recherche FNSP, Centre Emile Durkheim, Institut d'Etudes Politiques de Bordeaux, *Directeur de thèse*

M. François GONON,

Directeur de recherche émérite, Institut des Maladies Neurodégénératives, Université de Bordeaux, *Directeur de thèse*

Recherche biomédicale et journalisme en situation d'incertitude. Validité des résultats de la recherche biomédicale et couverture médiatique

De nombreux articles dans les journaux scientifiques font état du manque de reproductibilité des études biomédicales. Si cette « crise » semble bien connue des chercheurs, il apparaît que les médias, et donc le grand public, n'en ont pas conscience.

Le manque de reproductibilité n'est pas choquant en soi : la recherche est un processus cumulatif et les résultats initiaux sont par définition incertains. Nous avons analysé la prise en compte de cette incertitude par les médias. Nous avons constitué une base de données regroupant des résultats de la recherche biomédicale. Nous avons montré que plus de 60% des études initiales n'étaient pas reproduites. Nous avons également montré que la presse couvrait préférentiellement ces études initiales, peu reproduites, et n'informait que rarement le public des invalidations. D'autre part, les journalistes ne semblent pas tenir compte de cette incertitude : les articles de presse mentionnent rarement la nécessité de valider les résultats. Notre enquête auprès de 21 journalistes a révélé qu'ils n'avaient pas conscience du faible taux de validité des résultats initiaux.

Cette thèse discute de l'influence grandissante de facteurs extérieurs à l'activité scientifique dans le processus de production de connaissances. En particulier, la prise en compte par les chercheurs et les institutions scientifiques de critères d'intérêt médiatique pourrait influencer les stratégies de recherche et la fiabilité des résultats scientifiques. D'autre part, la détérioration des conditions de travail des journalistes et leur méconnaissance du fonctionnement de la recherche soulèvent des interrogations importantes sur la pertinence des informations présentées dans la presse et sur la qualité du débat public des questions de santé.

Mots-clés : reproductibilité - recherche biomédicale - incertitude – journalistes - sociologie des sciences - sciences de la communication

The reproducibility crisis in biomedical research: an analysis of the reliability of biomedical studies published in peer-reviewed journals and their media coverage

Science's reproducibility crisis is largely discussed in scientific journals editorials and most scientists are familiar with the problem. But journalists and therefore the public do not seem to be aware of the lack of reliability of a significant part of biomedical research.

This lack of reproducibility is to be expected: knowledge production is an incremental process and initial results are uncertain *per se*. Here we studied how the media presented this uncertainty when dealing with biomedical findings. We created a large database of biomedical studies. We assessed the validity of initial studies and showed that more than 60% were subsequently invalidated. We demonstrated that the press preferentially covered those initial results and scarcely mentioned their invalidation. Moreover, journalists do not seem to take into account the uncertainty of initial results: newspaper rarely specified that the results needed validation. Our survey of 21 journalists revealed that most of them were not aware of the lack of reproducibility of initial results.

Here we discuss the influence of extra-scientific factors on the production of scientific knowledge. We think that the scientific assessment process based on the number of paper published in high impact factor journals combined with the scientific institutions' orientation towards the media might undermine the reliability of scientific results, in academic publications as well as in the media. Indeed, journalists' working conditions are deteriorating and most do not seem to properly grasp how scientific facts are produced. This might be damaging for public trust in biomedical research and the debate about health-related issues.

Keywords: Reproducibility – Biomedical Research – Uncertainty – Journalists – science studies – science communication

Unité de recherche

Centre Emile Durkheim
Université de Bordeaux
Sciences Po Bordeaux
CNRS UMR 5116
11 allée Ausone
33607 Pessac cedex

Institut des Maladies Neurodégénératives
Université de Bordeaux
CNRS UMR 5293
Centre Broca Nouvelle-Aquitaine
146 rue Léo Saignat
33076 Bordeaux cedex
France

Remerciements

Ces remerciements sont pour moi l'occasion de témoigner ma reconnaissance à toutes les personnes qui m'ont soutenue et accompagnée lors de cette aventure d'entamer une nouvelle thèse à un âge plus avancé que la moyenne...

Je tiens en premier lieu à exprimer toute ma gratitude et ma reconnaissance à mes directeurs de thèse, Andy Smith et François Gonon. Plus particulièrement pour avoir toujours été très disponibles et pour leur soutien sans faille dans cette reconversion des sciences biologiques aux sciences sociales. Je les remercie également pour leurs critiques constructives, toutes les discussions stimulantes que nous avons pu avoir et enfin pour m'avoir donné la chance et les conditions idéales pour réaliser ce doctorat.

Je remercie également les membres du jury, Mme Isabelle Boutron, M. Tommaso Venturini et Mme Lucienne Letellier d'avoir accepté d'examiner et de lire ma thèse.

Mes remerciements vont également à l'Initiative d'excellence de l'Université de Bordeaux et au programme qui m'a financée pendant les trois années de mon doctorat et m'a permis de pouvoir assister à des conférences internationales et de publier en partenariat avec l'Université de Bristol (UK).

Je n'oublie pas les membres de l'équipe 4 de l'Institut des Maladies Neurodégénératives et en particulier Thomas Boraud pour m'avoir acceptée au sein de son équipe, pour son aide et les échanges précieux lors de la préparation et de la rédaction des publications et notamment sur les aspects statistiques.

Merci également aux autres doctorants de l'école doctorale SP2 que j'ai pu rencontré aux cours de séminaires ou journées de formation et avec lesquels nous avons échangé doutes et bonne humeur. Je pense en particulier à Camille, Caroline, Pauline, Eric et Vincent.

Bien évidemment, je n'aurais pas pu mener à bien cette thèse sans le soutien et les encouragements de ma famille et de mes proches. Merci à mes amis, mes parents, mes frères et sœurs et à mon mari Nicolas. Je dédie ce travail à nos deux enfants Joshua et Rebecca qui ont vu leur maman retourner « à l'école ».

Sommaire

UNITE DE RECHERCHE.....	3
REMERCIEMENTS	4
SOMMAIRE	5
PREAMBULE	9
CHAPITRE I.....	11
INTRODUCTION	13
<i>I- La crise de la reproductibilité</i>	18
1.1- Fraude, plagiat, falsification et pratiques de recherche discutables	20
1.2- « Publish or Perish » et fonctionnement de la recherche	22
<i>II- L'activité scientifique vue par les sciences sociales</i>	24
2.1- Ethos des sciences. Les normes Mertonniennes	25
2.2- Du mode I au mode II de production des connaissances ?	28
2.3- Reproductibilité et incertitude : éclairages de la sociologie des sciences	30
2.4- Les politiques de recherche et l'évaluation de la recherche.....	32
2.5- Sociologie de la promesse	34
<i>III- Sciences et Médias</i>	37
3.1- Couverture médiatique des sciences	39
3.2- Les relations chercheurs-journalistes	41
3.3- Le concept de <i>médialisation</i>	42
3.4- Les médias et l'incertitude	43
<i>IV- Notre thèse centrale</i>	45
CHAPITRE II.....	49
INCERTITUDE ET RECHERCHE - ANALYSE DE LA REPRODUCTIBILITE DES RESULTATS DE LA RECHERCHE BIOMEDICALE.....	51
<i>I- Méthodologie</i>	55
1.1- Sélection des méta-analyses	55
1.2- Sélection des études initiales et des études les plus larges	56
1.3- Extraction des données.....	56
1.4- Sélection de méta-analyses robustes	57
1.5- Analyse de la sensibilité et de la spécificité.....	57
<i>II- Résultats</i>	58
2.1- Caractéristiques des études incluses dans la base de données	58
2.2- Accord statistique entre les études initiales et les méta-analyses	59
2.3- Les tailles d'effet des études initiales sont elles exagérées ?	60
2.4- Taux de reproductibilité des études initiales.....	63
2.4.1- En fonction des différents domaines : psychiatrie, neurologie et maladies somatiques	63
2.4.2- En fonction du type d'association étudié.....	64
2.5- Influence de la taille des échantillons et des tailles d'effet sur la reproductibilité des études initiales	66
2.6- Taux de reproductibilité des études les plus larges.....	67
2.7- Influence du facteur d'impact du journal dans lequel l'étude initiale est publiée	68
2.8- Influence de l'année de publication de l'étude initiale sur le taux de reproductibilité	68
2.9- Evolution du nombre moyen de patients et des tailles d'effet moyennes des études initiales.....	69
<i>III- Apports et limites de cette étude</i>	70
3.1- Les apports	70
3.2- Les limites	73
<i>IV- Conclusion</i>	74
CHAPITRE III.....	79
INCERTITUDE ET MEDIAS - QUANTIFICATION ET ANALYSE DE LA COUVERTURE MEDIATIQUE DES RESULTATS DE LA RECHERCHE BIOMEDICALE	81
<i>I- Couverture médiatique des résultats de la recherche biomédicale</i>	83
1.1- Méthodologie	83
1.1.1- Sélection des études scientifiques.....	83
1.1.2- Influence du facteur d'impact des journaux scientifiques dans lequel les études ont été publiées	84
1.1.3- Identification des articles scientifiques couverts par la presse	85
1.1.4- Reproductibilité des études scientifiques couvertes par la presse	86
1.2- Résultats	86
1.2.1- Etudes scientifiques couvertes par la presse	86
1.2.2- Influence du domaine biomédical et du type d'association analysé	87
1.2.3- Couverture médiatique des études initiales et des études ultérieures.....	88
1.2.4- Couverture préférentielle des résultats positifs.....	89
1.2.5- Reproductibilité des études scientifiques couvertes par la presse.	90
1.2.6- La presse couvre-t-elle les résultats d'études contredisant ceux d'études qu'elle a précédemment couvertes ?.....	92

1.3- Apports et limites de cette étude	93
1.3.1- les apports	93
1.3.2- Les limites	94
1.4- Conclusion	95
II- Présentation des résultats de la recherche biomédicale dans la presse grand public	96
2.1- Méthodologie	97
2.1.1- Sélection des articles de presse	97
2.1.2- Analyse du contenu des articles de presse	98
2.2- Résultats	100
2.2.1- Validité des codages	100
2.2.2- Description des articles de presse couvrant les études initiales	101
2.2.3- Analyse des corrélations entre la description d'une étude comme initiale, la mention de l'incertitude ou de la fiabilité des résultats	102
2.2.4- Influence des scientifiques sur la mention de la nécessité de reproduire les résultats	102
2.2.5- L'exagération des titres est inversement corrélée à la mention de la nécessité de reproduire les résultats	103
2.2.6- Influence du taux de reproductibilité	103
2.2.7- Influence de l'année de publication de l'étude scientifique	104
2.3- Apports et limites de l'analyse de contenu des articles de presse couvrant des études initiales	105
2.3.1- L'incertitude scientifique dans la presse	105
2.3.2- Limites de l'étude	106
2.4- Conclusion	107
CHAPITRE IV	111
JOURNALISTES SCIENTIFIQUES - PRATIQUES ET REPRESENTATIONS	113
I- Journalisme scientifique : approche sociologique	114
1.1- Caractéristiques du journaliste scientifique	114
1.2- Les normes journalistiques	117
1.3- <i>Gatekeeping Theory</i> : Facteurs influençant le choix des sujets à couvrir	119
1.3.1- Facteurs individuels	120
1.3.2- Routines et pratiques des journalistes	121
1.3.3- Facteurs organisationnels	122
1.3.4- Facteurs externes	123
1.3.5- Facteurs sociétaux	123
1.4- Relations aux sources : interactions entre journalistes scientifiques et chercheurs	124
II- Une enquête auprès des journalistes scientifiques sur leurs pratiques et leur position face à l'incertitude scientifique	128
2.1- Présentation de l'enquête	128
2.2- Sociographie des journalistes interviewés	129
2.3- Représentations des choix du sujet et des sources	131
2.3.1- Choix des sujets	131
2.3.2- Sources utilisées par les journalistes	133
2.4- Préparation du document	137
2.4.1- Vérification auprès de l'article primaire	137
2.4.2- Vérification des informations	138
2.4.3- Ecriture et remaniements par la rédaction	139
2.4.4- Choix du titre	142
2.5- Interactions avec les scientifiques	144
2.5.1- interview des auteurs des études scientifiques	144
2.5.2- Rapport avec les chercheurs	145
2.6- Opinion sur la reproductibilité des sciences	149
2.7- Réflexions sur leur rôle	154
III- Conclusion	155
3.1- L'enquête et ses limites	156
3.2- Influences des différents niveaux du <i>Gatekeeping</i>	157
3.2.1- Facteurs individuels	158
3.2.2- Le poids des routines	159
3.2.3- Le poids de la hiérarchie	160
3.3- Méconnaissance du fonctionnement de la recherche	161
3.4- l'illusion du 4 ^{ème} pouvoir ?	162
CHAPITRE V	165
CONCLUSION	167
I- Crise de la reproductibilité et journaux scientifiques à fort facteur d'impact	169
II- Politique d'évaluation de la recherche, financements et journaux à fort facteur d'impact	170
III- La visibilité médiatique et journaux à fort facteur d'impact	173
IV- Vers un nouveau regard sur les rapports sciences-médias	177

BIBLIOGRAPHIE	181
ANNEXES	197

Préambule

De nombreux articles dans les journaux scientifiques évoquent de plus en plus souvent les problèmes de reproductibilité des résultats scientifiques à tel point qu'on parle de « crise de la reproductibilité ». Selon leurs auteurs, cette crise sape les fondations de la recherche scientifique et questionne son intégrité.

Ce constat d'une mauvaise reproductibilité des recherches recouvre en fait plusieurs problèmes : la fabrication de résultats (fraude), la falsification des données et les pratiques de recherche discutables. Ces problèmes ont été discutés et étudiés par des disciplines ou courants de recherche telles que l'éthique et l'intégrité des sciences, la science politique, la sociologie des sciences et les sciences de la communication. Par contre ce qui n'est quasiment jamais pris en compte par les commentateurs de « la crise de la reproductibilité » c'est que la recherche scientifique est par nature incertaine.

En effet, la recherche scientifique est un processus cumulatif : elle part de résultats prometteurs mais incertains pour arriver à un consensus après réplique des observations par les pairs. Le fait que de nombreux travaux initiaux ne soient pas confirmés ne contredit donc pas les normes traditionnelles de la recherche scientifique. Le point de départ de cette thèse est que l'incertitude inhérente à la recherche en train de se faire n'est pas prise en compte dans les interactions entre recherche et société. Même au niveau académique, les études en sociologie des sciences ou en sciences de la communication s'intéressent à l'incertitude dans le cas des controverses scientifiques lorsque le consensus scientifique n'est pas atteint ou lorsque les implications environnementales ou sanitaires des nouvelles technologies ne sont pas connues. Nous n'avons pas trouvé dans la littérature d'études qui posent la question de l'incertitude liée aux résultats initiaux et à leur utilisation dans les médias ou pour la décision publique.

A l'heure où l'on parle de « démocratie technique », du « contrat social entre science et société » et de la participation du public au débat scientifique, il nous a donc semblé pertinent d'examiner la présentation des résultats initiaux des sciences biomédicales dans les médias, plus particulièrement dans la presse écrite. Plus exactement, en se basant sur l'étude de la couverture médiatique d'études scientifiques dont la validité est connue, cette thèse essaie de comprendre le rôle des scientifiques, de leurs institutions et des journalistes dans la présentation des sciences dans les médias. Nous tenterons de saisir, en utilisant des concepts empruntés à la sociologie des sciences, à la sociologie du journalisme et aux sciences de la communication, les jeux de pouvoir qui opposent ces différents acteurs dans la mise à l'agenda¹ de problèmes publics.

¹ Mettre à l'agenda ou *agenda setting* : « Désigne la capacité qu'ont les médias, par la sélection des nouvelles, de produire une hiérarchisation de l'information et d'exercer ainsi une influence qui n'est pas tant de modeler des comportements que de définir des thèmes dignes de l'attention collective » (Neveu, 2009, p. 85).

CHAPITRE I

INTRODUCTION

Selon le rapport rédigé pour l'Eurobaromètre 2007, « les découvertes scientifiques et leurs applications pratiques participent au développement industriel et à la promotion sociale. La recherche, l'innovation et les nouvelles technologies influencent les dimensions économique, sociale, politique et parfois même éthique des sociétés modernes ». L'enquête commanditée par la direction générale de la recherche de la Commission européenne avait ainsi pour but de « découvrir ce qui incitait les citoyens européens à s'engager davantage dans les sciences, la recherche et l'innovation » et « d'explorer la manière dont les médias pourraient contribuer à encourager cet engagement des citoyens européens » (Eurobaromètre 2007). Les liens entre recherche scientifique, médias, acceptation publique et décision politique sont dès lors mis en lumière. Il devient impératif que les scientifiques aient la confiance du public et que les résultats de la recherche soient les plus fiables et robustes possibles. C'est pourquoi de nombreux chercheurs, notamment en sciences biomédicales, s'inquiètent de la « crise de la reproductibilité » qui frappe la recherche biomédicale. Celle-ci est encore peu relayée par la presse (The Economist 2013a, 2013b) mais préoccupe néanmoins les revues scientifiques, les agences de financement et les institutions scientifiques (Baker 2016b; Collins et al. 2014). Ainsi, au Royaume-Uni, en octobre 2015, *the Academy of Medical Sciences* en collaboration avec les principales institutions de recherche et de financement en sciences biologiques et médicales (BBSRC, MRC et Wellcome Trust²), a publié un rapport analysant certaines causes de la non reproductibilité des sciences biomédicales et a proposé des solutions pour palier à ce problème (Academy of Medical Sciences et al. 2015).

De plus, les grandes crises sanitaires et environnementales exposées dans les médias (la maladie de la vache folle, le sang contaminé, les OGM (Organismes Génétiquement Modifiés) et les perturbateurs endocriniens pour en citer quelques unes) ont ébranlé la confiance du public et ont conduit les chercheurs en sciences sociales à s'interroger sur la légitimité des sciences dans la prise de décision, sur ses modes de production et sur sa transmission au public. Aux Etats-Unis, les *National Academies of Sciences, Engineering, and Medicine* ont publié récemment un ouvrage intitulé « *Communicating science effectively : a Research Agenda* » qui répertorie les approches efficaces de communication des sciences ainsi que des pistes de recherche pour les sciences de la communication (Academy of Medical

² BBSRC : Biotechnology and Biological Science Research Council. MRC : Medical Research Council, Wellcome Trust : association caritative britannique qui finance des projets de recherche biomédicaux

Sciences et al. 2015).

En France aussi des comités de réflexion sur l'intégrité de la recherche se sont développés et ont organisé un premier colloque qui s'est déroulé à Bordeaux en janvier 2016 sous le titre « L'intégrité scientifique, parlons-en ! »³. Sur le plan international, l'implication des différentes institutions au niveau mondial est illustrée par les conférences internationales sur l'intégrité scientifique (WCRI : *World Conference on Research Integrity*) dont la cinquième s'est déroulée à Amsterdam en mai 2017. Cette conférence a rassemblé des acteurs scientifiques de différentes disciplines : les sciences biologiques et médicales, les sciences de la communication, les sciences politiques, les statistiques. Parmi les 800 participants on comptait aussi des éditeurs de grands journaux scientifiques (i.e. *Nature*, *Science*), des responsables d'agences de financement (i.e. National Institutes of Health, European Research Council) et enfin des décideurs politiques (ministres de la santé, des sciences et technologies, membres du conseil européen ainsi que des présidents d'université).

L'organisation de ces colloques et conférences ou encore la rédaction de ces ouvrages souligne bien l'importance accordée à l'activité scientifique et à sa légitimité. Bien entendu, de nombreuses équipes de recherche tentent donc d'aborder les problèmes d'intégrité scientifique en réfléchissant à ses causes, ses implications ou plus largement à la place de la science dans la société.

Pour comprendre les problèmes de reproductibilité des résultats scientifiques, il convient de rappeler que la science a pour but de produire des connaissances. Pour les sciences de la vie, la recherche procède selon une méthode hypothéico-déductive : les chercheurs partent d'une hypothèse qu'ils testent avec des expériences. Les résultats de ces expériences sont alors analysés et proposés pour publication aux pairs. Ceux-ci jugent la qualité du travail et des observations et acceptent ou non leur publication. Les résultats sont alors accessibles à la communauté scientifique qui peut dès lors les valider et ainsi produire des connaissances solides grâce à la réplication des expériences. Un résultat initial est donc par définition incertain et en attente de validation. Or cette incertitude liée aux résultats initiaux semble être méconnue des sociologues, chercheurs en science de la communication, journalistes et décideurs comme nous le décrivons au cours de cette thèse.

Ainsi, dans leur ouvrage « Agir dans un monde incertain » les auteurs Michel Callon, Pierre Lascoumes et Yannick Barthe décrivent l'incertitude liée aux sciences et techniques, notamment dans le domaine de la santé et de l'environnement (Callon et al. 2014). Tout d'abord, ils distinguent explicitement l'incertitude du risque : le risque existe quand les

³ www.u-bordeaux.fr/Actualites/De-la-recherche/L-integrite-scientifique-parlons-en

conséquences potentielles sont correctement identifiées sans savoir si elles vont se produire. Dans le cas des risques, il est alors possible, pour les décideurs politiques, de prendre des décisions basées sur la connaissance des différents scénarios possibles. L'incertitude est donc différente du risque : elle intervient lorsque la science ne permet pas d'envisager rationnellement tous les « états du monde possibles ». Les auteurs donnent alors différentes définitions de l'incertitude : « entre la sombre ignorance et l'impeccable connaissance des états du monde possibles, s'étend un vaste espace » (Callon et al. 2014). Ils expliquent l'importance de l'utilisation des controverses pour pouvoir appréhender et « repérer les contours » des différentes formes d'incertitudes. Ils nous engagent alors à distinguer des incertitudes plus ou moins radicales. Premièrement, « les risques de développement » liés à la commercialisation de substances qui peuvent avoir des effets non immédiatement mesurables (par exemple les effets du distilbène sur la descendance des femmes qui ont été traitées par cette substance ou encore l'affaire du sang contaminé). Dans ce cas, l'incertitude ne peut être mesurée qu'*a posteriori*. Deuxièmement, le « danger potentiel plausible » pour les cas où des personnes ou des milieux sont affectés par des dommages caractérisés mais dont les causes sont inconnues. C'est ce que les auteurs définissent comme l'« ère du soupçon » et mentionnent les controverses liées aux effets cancérigènes des radiations émises par les téléphones portables ou des substances radioactives présentes dans l'environnement. Enfin, les auteurs définissent l'incertitude liée aux présomptions : les phénomènes sont établis et robustes et personne ne les conteste. L'incertitude repose sur la chaîne causale comme par exemple dans le cas de l'ESB (Encéphalite Spongiforme Bovine). Les vaches sont porteuses d'un agent infectieux encore non identifié mais qui existe et il est probable que cet agent soit transmissible à l'homme (Callon et al. 2014).

On voit bien ici que l'incertitude décrite par ces auteurs est celle liée à la science émergente et au fait que, compte tenu de l'état actuel des connaissances, une décision peut être difficile à prendre (réchauffement climatique, enfouissement des déchets radioactifs, utilisation des OGM). L'incertitude liée à la construction de faits scientifiques, qui ne sont pas connus avec certitude parce que les résultats sont initiaux, n'est pas envisagée. Prenons l'exemple de la découverte d'un polymorphisme génétique impliqué dans une maladie sur une certaine population par une équipe de scientifiques. Cette découverte n'est pas ouverte à la controverse au sens sociotechnique. Il n'y a pas d'incertitude quant à ses conséquences ou ces effets à plus ou moins long terme. L'incertitude est ici intrinsèque au résultat : il faut le reproduire pour le rendre avéré, sur d'autres populations. Une seule étude réalisée sur un sujet ne constitue pas un fait scientifique. C'est sur ce type d'incertitude, celle liée aux résultats initiaux, que porte cette thèse. Expliquer au public - et notamment aux décideurs - ce type

d'incertitude est indispensable pour que ceux-ci n'attribuent pas une confiance exagérée aux résultats scientifiques et qu'ils puissent prendre des décisions de manière aussi justifiée que possible (Fischhoff et al. 2014).

On voit dès lors qu'incertitude et intégrité sont liées : les scientifiques ne doivent pas présenter leurs résultats comme plus robustes qu'ils ne le sont réellement, en particulier lorsque ceux-ci sont utilisés pour prendre des décisions publiques. En effet, ils ont une responsabilité vis à vis du public : celui-ci finance et utilise les résultats de la recherche. Si ces résultats sont floués, la confiance du public est ébranlée. Et il est important de conserver la confiance de la société et d'obtenir son appui, ainsi que celui des décideurs, pour pouvoir utiliser au mieux les données scientifiques et « agir dans un monde incertain » (Callon et al. 2014).

C'est pourquoi de nombreuses recherches sont conduites sur les sujets d'intégrité scientifique, de reproductibilité des observations scientifiques, de qualité et forme de la communication des résultats de la recherche. Ainsi, les sciences biologiques et médicales étudient la reproductibilité des résultats scientifiques d'un point de vue quantitatif (Begley et al. 2012; Prinz et al. 2011). Elles essaient d'identifier les causes, d'expliquer les problèmes de reproductibilité (Ioannidis 2005, 2008; Munafo et al. 2014) et les façons d'y remédier (Chalmers et al. 2014; Macleod et al. 2014). Les sciences de la communication s'intéressent elles à la présentation des sciences dans les médias (de Semir et al. 1998; Pellechia 1997), aux interactions entre scientifiques et journalistes (Peters 1995), à la façon dont les sciences sont communiquées (Trench 2008). Elles examinent en particulier les problèmes de présentation des risques (Friedman et al. 1999), de l'incertitude liée aux sciences émergentes (biotechnologies, réchauffement climatique), des fraudes et de leurs conséquences sur le public. La sociologie (politique) des sciences dépasse ces cadres d'analyse quantitatifs et descriptifs et s'intéresse à la science en tant qu'activité sociale (Bucchi 2015), à sa légitimité dans la prise de décision (Wynne 1996) et aux politiques d'évaluation et de recherche (Fouquet et al. 2010). Lorsqu'elle traite de l'incertitude, c'est de celle liée aux controverses scientifiques ; qu'elles soient internes au champ scientifique en raison de désaccord entre les chercheurs ou qu'elles soient liées au manque de connaissances ou à l'incapacité des sciences seules à guider la décision.

Il semble donc que le premier type d'incertitude, l'incertitude inhérente aux résultats initiaux, ne soit pas prise en compte. C'est ce que nous nous proposons de faire dans le cadre de cette thèse. Plus précisément, nous avons analysé la reproductibilité et la couverture

médiatique des sciences biomédicales. En effet, celles-ci suscitent souvent l'intérêt des médias (Bauer 1998). Elles représentent un large pourcentage des études scientifiques couvertes par la presse (Suleski et al. 2010). Le public est donc régulièrement informé des dernières avancées en matière de diagnostic, de traitement mais aussi en matière de risque et de prévention du risque. C'est sur ce dernier élément que se focalise cette thèse. Nous avons choisi d'analyser la couverture médiatique des résultats d'observations scientifiques qui étudiaient l'association entre différents facteurs de risque (environnementaux, biochimiques, génétiques) et des pathologies. Ce type de recherche est en effet souvent repris par les médias du fait de son implication directe pour le lecteur.

Un autre élément vient légitimer le choix de sciences biomédicales comme objet de recherche. En effet, dans leur étude sur la cartographie des différents domaines de recherche scientifique (Rafols et al. 2010), les sciences biomédicales représentent un pôle majeur de l'activité de recherche (Figure 1).

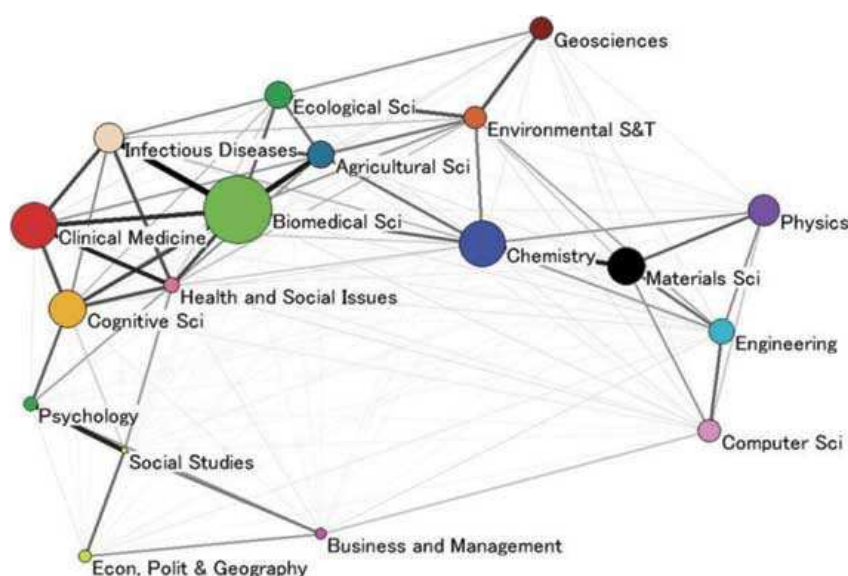


Figure 1 : La structure des sciences. La taille de chaque pastille est proportionnelle au nombre de citations produites. D'après Rafols et al. 2010 (Rafols et al. 2010).

Dans les paragraphes qui suivent, nous décrivons les apports et cadres fournis par les différentes disciplines pour comprendre et étudier l'incertitude liée aux résultats initiaux de la recherche et sa présentation au public par l'intermédiaire des médias. Dans une première partie, nous détaillons ce que les auteurs en sciences biologiques et médicales entendent par « la crise de la reproductibilité », ses causes et les moyens qu'ils proposent pour y remédier. Nous évoquons ainsi le lien qui existe entre incertitude et reproductibilité. Dans une deuxième partie, en utilisant des éclairages fournis par la sociologie et la science politique, nous définissons l'activité scientifique et présentons la définition que ces disciplines font de l'incertitude. Ceci nous permet de souligner la prégnance des normes mertonniennes dans

l'activité menée par les scientifiques en recherche biomédicale et les conflits qui en résultent. Dans une troisième partie, nous présentons les études réalisées en sciences de la communication sur la présentation des sciences dans les médias et introduisons le concept de *médialisation*. Enfin, nous explicitons les différents apports de ces disciplines et présentons l'objectif de la thèse au travers de l'analyse la présentation de l'incertitude liée aux résultats initiaux de la recherche biomédicale dans les publications scientifiques mais aussi dans les médias.

I- La crise de la reproductibilité

Nous expliquons ici ce que les auteurs en sciences biologiques et médicales entendent lorsqu'ils évoquent « la crise de la reproductibilité ». Nous présentons aussi quelques conclusions d'observations réalisées sur la reproductibilité de résultats obtenus dans des disciplines telles que la psychologie, les sciences biomédicales, ou encore dans le cas d'études cliniques et précliniques. Enfin, nous décrivons les causes qui sont avancées pour expliquer les problèmes de reproductibilité et l'influence du mode de financement et fonctionnement de la recherche.

XXV

« Toutes les expériences qui seront rapportées par quelque Académicien, seront vérifiées par lui dans les Assemblées, s'il est possible, ou du moins elles le seront en particulier en présence de quelques académiciens. »

XXIX

« L'académie fera de nouveau les Expériences considérables qui se sont faites partout ailleurs, on marquera dans les registres la conformité ou la différence des siennes à celles dont il était question. »

Articles XXV et XXIX Charte de l'académie royale des sciences (1699)

Dès le début de l'institutionnalisation des sciences, l'importance de la reproductibilité d'un résultat est soulignée. Elle en est une caractéristique essentielle. La méthode scientifique part d'une hypothèse que le chercheur teste par des expériences. Les résultats obtenus sont alors décrits dans un texte examiné par les pairs (les "*reviewers*") qui recommandent ou non sa publication. Si le texte est publié, les expériences peuvent alors être reproduites par la communauté scientifique, et ainsi être confirmées (ou non). Seules des expériences

confirmées peuvent aboutir à des connaissances fiables. La qualité du savoir scientifique est donc assurée par ce système d'autocorrection mis en place par la communauté scientifique elle-même. Ce système est donc en quelque sorte garant de l'intégrité scientifique : il permet d'identifier les fraudes, plagiat et autres pratiques discutables et corrige les résultats erronés.

Cependant le système semble débordé, il y a un « trop-plein » de productions scientifiques (Siebert et al. 2015) depuis que nous sommes passés d'une « *little science* » à une « *big science* » pour reprendre Derek de Solla Price. Dans son ouvrage de 1986, il décrit la croissance exponentielle de la production scientifique, du nombre de chercheurs et des financements attribués à la recherche (Price 1986). Il expose également la possibilité de « faire une science de la science » ou scientométrie pour mesurer et étudier la production scientifique, posant ainsi les jalons de l'évaluation scientifique.

Récemment, le journal *Nature* a publié les résultats d'une enquête menée auprès de 1576 scientifiques : pour 90% d'entre eux il existe une crise de la reproductibilité en sciences (Baker 2016b). De plus, entre 60 et 80% des scientifiques interrogés reconnaissent qu'ils n'ont pas réussi à reproduire des résultats déjà publiés (Baker 2016b).

Le constat n'est pas nouveau. Depuis une dizaine d'années, les journaux scientifiques, notamment dans les domaines de la biomédecine et de la psychologie, font état de ces problèmes de reproductibilité (Casadevall et al. 2010; Ioannidis 2005; Ioannidis et al. 2014; 2015). Cette crise touche de nombreuses disciplines : la recherche préclinique (Begley et al. 2012; Prinz et al. 2011) où les auteurs n'ont pas pu reproduire l'immense majorité des résultats (90% et 70% respectivement), les résultats de puces à ADN (Ioannidis et al. 2009), ou encore les études d'association entre des pathologies et des marqueurs biologiques (Ioannidis et al. 2011c). Cette crise constitue aussi un obstacle pour le transfert des résultats de laboratoire en essais cliniques (Mobley et al. 2013).

Plusieurs causes ont été mises en avant pour expliquer ces problèmes de reproductibilité : la trop faible puissance statistique⁴ des études (Button et al. 2013), le fait de changer son hypothèse après que les résultats aient été obtenus (Kerr 1998), la mauvaise utilisation des statistiques, en particulier le seuil de $p < 0,05$ ou encore le « *p hacking* » c'est à dire la sélection des données de manière à obtenir un résultat statistiquement significatif (Kass et al. 2016; Nuzzo 2014; Weissgerber et al. 2015). Beaucoup s'inquiètent de la prépondérance de l'utilisation du système de l'hypothèse nulle en sciences biomédicales. Dans ce modèle, le scientifique s'efforce de réfuter une hypothèse nulle (par exemple, il n'y a pas de corrélation

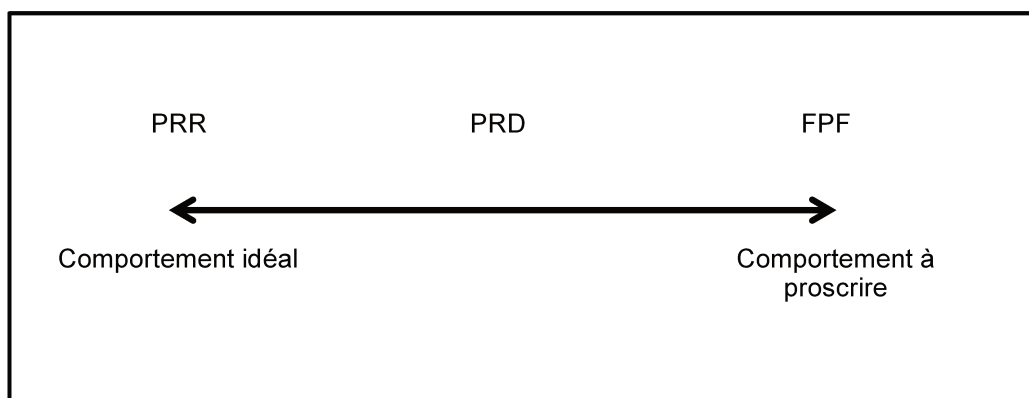
⁴ La puissance statistique d'une étude fait référence à la capacité de cette étude à pouvoir détecter un effet si celui-ci existe. Elle est liée à la taille de l'échantillon et à la taille d'effet (la force de la relation qui existe entre les 2 paramètres).

entre fumer et développer un cancer du poumon) en utilisant comme seuil $p=0,05$ (Baker 2016; Wasserstein et al. 2016). Ils calculent la probabilité d'obtenir les résultats qu'ils observent si l'hypothèse nulle était vraie⁵. Des journaux comme le journal *Epidemiology* ont ainsi critiqué l'utilisation du $p<0,05$ (Nuzzo 2014). D'autres prônent l'utilisation des statistiques bayésiennes qui proposent de « penser les probabilités comme la plausibilité qu'un fait se produise plutôt que comme la fréquence potentielle de survenue de ce fait » (Nuzzo 2014). Leur utilisation aurait pour avantage, pour les scientifiques, de pouvoir tenir compte de ce qu'ils connaissent de leur projet plutôt que de partir de l'hypothèse nulle.

Plusieurs auteurs ont montré que ces pratiques conduisaient à la publication d'un excès de résultats positifs (statistiquement significatifs) et donc souvent réfutés par les études ultérieures (Fanelli et al. 2017; Ioannidis 2005b, 2008, 2011b). Mais comment expliquer ces phénomènes ? Pourquoi le système d'évaluation par les pairs ne garantit-il pas de la reproductibilité ?

1.1- Fraude, plagiat, falsification et pratiques de recherche discutables

Plusieurs auteurs se sont intéressés au comportement des chercheurs lorsqu'ils pratiquent la recherche notamment au travers d'interviews (de Vries et al. 2006; Fanelli 2009; Martinson et al. 2005; Steneck 2006). Ainsi dans son article, Steneck présente les trois catégories de comportement qui peuvent décrire l'activité des chercheurs. Il les représente le long d'un axe horizontal avec à gauche un comportement idéal et à droite un comportement à proscrire (Figure 2). On trouve, en regard du comportement idéal, les pratiques de recherche responsables (PRR), et à son opposé les phénomènes de fraude, plagiat et falsification (FPF) avec, entre les deux, les pratiques de recherche discutables (PRD).



⁵ Donc avec un $p=0,05$, ils acceptent la possibilité de se tromper c'est à dire de rejeter à tort l'hypothèse nulle dans 1 cas sur 20. En choisissant un tel seuil, on voit bien qu'il est normal que certains résultats ne soient pas reproductibles. C'est pourquoi, certains préconisent l'abandon de son utilisation ou d'augmenter les seuils à $p<0,001$.

Figure 2 : Différentes pratiques de la recherche. PRR : Pratiques de recherche responsable, PRD : pratiques de recherche discutables, FPF : Fraude, Plagiat et Falsification. D’après Steneck (Steneck 2006).

Ces pratiques de recherche discutables se situent dans cette zone grise, où le chercheur ne commet pas de faute dite grave et facilement repérable. Elles recouvrent un grand nombre de phénomènes différents dont certains sont présentés dans le tableau 1 (Ioannidis et al. 2014; Steneck 2006).

Tableau 1 : Exemples de pratiques de recherche discutables

Non déclaration des conflits d’intérêt
Découpage des résultats en plusieurs publications (« <i>salami slicing</i> »)
Biais de citation : citation uniquement les études qui confirment les hypothèses
Biais de publication : présente uniquement des résultats positifs
Omission ou sélection des données
Mauvaise utilisation des statistiques (intentionnelle ou non)
Mauvaise conception des études (petit échantillon, faible puissance statistique)
Exagération des résultats
Co-signature d’auteurs qui n’ont peu ou pas participé à la recherche

En 1992, les *National Academies of Science* ont défini ces PRD comme « des actions qui s’opposent aux valeurs traditionnelles de l’entreprise de recherche et qui pourraient être problématiques pour son fonctionnement. » Elles les ont différencié des FPF parce « qu’elles ne nuisent pas directement à l’intégrité du fonctionnement de la recherche » (Steneck 2006). Ce positionnement est discutable dans la mesure où leur détection est plus difficile et que les nombreuses publications qui analysent la mauvaise reproductibilité des résultats de la recherche scientifique montrent que ce sont justement ces pratiques de recherche (mauvaise conception des études, faible puissance statistique etc.) qui en sont souvent la raison.

Plusieurs enquêtes ont été réalisées auprès des chercheurs. Tous soulignent que les fraudes, falsification et plagiat sont des problèmes rares ; les chercheurs s’inquiètent d’avantage des pratiques de recherche discutables (Siebert et al. 2015). Dans l’enquête réalisée par Martinson en 2005 sur plus de 3000 chercheurs américains, 33% d’entre eux ont admis avoir utilisé des pratiques de recherche discutables au cours des 3 dernières années (Martinson et al. 2005). Dans une méta-analyse de 2009, Fannelli compile les différentes études qui ont été réalisées auprès des scientifiques pour analyser leurs comportements et pratiques (Fannelli 2009). Les résultats sont révélateurs : 2% des scientifiques reconnaissent avoir commis au moins une fois des fautes graves telles que la fabrication et la falsification des données et 33% reconnaissent des pratiques discutables. Quand ils sont interrogés sur les comportements de leurs collègues, les pourcentages passent alors à 14% pour les fautes

graves et 72% pour les pratiques discutables. Il semble donc que malheureusement, les pratiques de recherche discutables soient monnaie courante et qu'elles pourraient expliquer l'essentiel des problèmes de non-reproductibilité. C'est pourquoi pour bon nombre de chercheurs les pratiques discutables sont les plus dommageables pour la recherche et ce sont contre elles qu'il faut mettre en place des mesures (de Vries et al. 2006; Smaldino et al. 2016). Selon Fanelli (Fanelli 2009), ces « tricheries » qu'elles soient fraudes caractérisées, falsifications ou pratiques discutables sont commises ultimement dans le but de tromper et de « donner à une observation ordinaire l'apparence et les caractéristiques d'une observation de la plus haute qualité » (Babbage 1830). La non-reproductibilité des résultats de recherche implique donc indirectement l'éthique des chercheurs et donc l'intégrité de la recherche.

1.2- « *Publish or Perish* » et fonctionnement de la recherche

« *Faire de la recherche c'est un triathlon avec trois épreuves d'endurance : la course à l'innovation, la course aux financements et la course à la publication* ». Claude Huriet (Président du MURS-IS⁶)

Les publications qui font référence à la crise de la reproductibilité expliquent toutes que la majorité de ces phénomènes peut être expliquée par le mode de fonctionnement et d'évaluation de la recherche : pour réussir, c'est à dire obtenir un poste de recherche académique et des financements, il faut publier (Lawrence 2003; Young et al. 2008). Les critiques sont acerbes (Lawrence 2007; Smaldino et al. 2016). Il ne faut pas seulement publier, il faut publier beaucoup et dans des journaux prestigieux. En effet, la compétition importante qui existe entre les scientifiques pour l'obtention de financements et de postes, combinée aux indices bibliométriques qui permettent l'évaluation quantitative des chercheurs, favorisent cette pression qui pousse les scientifiques à poursuivre des recherches « publiables » (Fanelli 2010). Le nombre d'articles publiés explose, et on assiste à un accroissement sans précédent du nombre de nouveaux journaux. Ainsi le nombre de publications augmenterait de 3,5% par an et le nombre de journaux de 3% (Bucchi 2015).

Si l'on considère les critères d'évaluation valorisés par les journaux scientifiques et des agences de financement, la recherche se résumerait maintenant à produire et publier des découvertes (Ioannidis 2013). Ceci conduit les scientifiques à ne publier que des résultats positifs, à délaisser les résultats négatifs (Couzin-Frankel 2013; Schooler 2011) et à ne pas tenter de reproduire les expériences. De fait une étude a montré que, dans 70 à 90% des cas (en fonction des disciplines), les résultats publiés confirmaient l'hypothèse testée (Yong

⁶ MURS-IS : Mouvement Universel pour la Responsabilité Scientifique-Intégrité Scientifique.

2012). Il semble pourtant peu probable que toutes les hypothèses testées soient valides. Plusieurs raisons peuvent expliquer cet excès : les auteurs ont pu reformuler leur hypothèse après analyse des résultats de manière à obtenir un résultat positif, réaliser une analyse sélective des données ou encore ne pas essayer de publier des résultats négatifs. Quelle qu'en soit la raison, la présentation des résultats de la recherche est biaisée et ce dans toutes les disciplines (Yong 2012).

Le fait de favoriser la publication des résultats positifs statistiquement significatifs conduit également à des distorsions dans l'écriture des publications. Ainsi les résumés et conclusions tendent à accentuer les résultats positifs, amoindrir les résultats négatifs. Des analyses d'articles réalisées dans les cas des essais cliniques randomisés négatifs ou dans des publications en neurosciences, ont montré l'utilisation de telles exagérations dans les résumés et les conclusions (Boutron et al. 2010; Gonon et al. 2011). Ces exagérations se retrouvent aussi au niveau du vocabulaire. Ainsi, des auteurs ont analysé l'évolution de l'utilisation des termes positifs dans tous les résumés indexés par Pubmed⁷ publiés entre 1974 et 2014 (Vinkers et al. 2015). Ils ont suivi l'évolution de mots positifs tels que « nouveau », « robuste », « innovant » ou « sans précédent » et ont montré que l'utilisation de ces mots avait augmenté de plus de 2500%. En comparaison, les mots neutres et négatifs n'avaient pas ou peu augmenté (Vinkers et al. 2015).

De plus, pour publier beaucoup, les chercheurs ont tendance à « découper » leurs résultats pour proposer des micro-études à plusieurs journaux et ainsi maximiser, en terme de nombre de publications, la rentabilité de leurs travaux (Bertamini et al. 2012). On voit bien que la qualité de la recherche est ainsi biaisée : un résultat publiable n'implique plus que ce résultat soit fiable (Nosek et al. 2012) et s'il n'y a aucune incitation, pour que les chercheurs tentent de reproduire les résultats, il est donc inévitable que la capacité des sciences à s'auto-corriger ne soit plus possible. Cette course à la publication est aussi néfaste pour la créativité des chercheurs. Elle peut les empêcher de poursuivre des projets plus compliqués et longs à réaliser puisqu'alors les temps pour publier seront allongés et qu'il leur sera donc plus difficile d'obtenir des financements (Alberts 2013; Ioannidis 2011).

Par ailleurs, les conséquences de la piètre reproductibilité des études biomédicales sont encore aggravées par les biais de citation. Ainsi, certaines études sont toujours citées comme probantes par des auteurs qui semblent ignorer qu'elles ont été invalidées (Tatsioni et al. 2007). La citation sélective des études en accord avec les hypothèses des auteurs conduit à

⁷ <https://www.ncbi.nlm.nih.gov/pubmed>

des dogmes infondés qui perdurent dans la littérature scientifique (Greenberg 2009). Enfin, en raison des biais de publication et de la visibilité des journaux à fort facteur d'impact, Begley et al. ont montré que des études précliniques non-reproductibles mais publiées dans des journaux prestigieux sont autant citées que celles qui sont reproduites (Begley et al. 2012). Elles « polluent » la discipline, engendrent de multiples publications secondaires basées sur une observation initiale erronée allant jusqu'à entraîner la mise en place d'essais cliniques dangereux pour les patients (Begley et al. 2012). On voit dès lors l'importance des journaux scientifiques à fort facteur d'impact pour la visibilité des recherches scientifiques.

Toutes ces observations, explications et recommandations concernant la médiocre réplication des résultats de la recherche biomédicale n'épuisent pas la question. On connaît mal l'entendue de ce phénomène, son évolution temporelle et ses conséquences. Le pourcentage de résultats non reproduits est-il plus grand dans certains domaines des sciences biomédicales que dans d'autres ? Est-il plus grand maintenant qu'il y a trente ou quarante ans ? Ce phénomène mine-t-il vraiment la confiance du public et le soutien à la recherche ?

Enfin, quelques scientifiques font remarquer que toutes ces recommandations contraignantes ont un coût. Par exemple la recommandation omniprésente de multiplier les observations pour augmenter la puissance statistique d'une étude préliminaire risque de favoriser les laboratoires et les universités les plus riches. Cette recommandation se heurte aussi à des exigences éthiques, en particulier pour les expériences chez l'être humain. De plus, certains scientifiques font remarquer que l'on ne peut pas et que l'on ne devrait pas s'attendre à une reproductibilité de 100% des études initiales. Il est normal pour la science et la construction des connaissances de procéder par tâtonnements et vérifications, la science est un processus cumulatif et aucune étude initiale sur un sujet ne saurait donner une réponse robuste et définitive à une question (Open Science Collaboration 2015). Malheureusement, cette remarque est très rarement exprimée, tant dans les journaux scientifiques que dans la presse grand-public.

II- L'activité scientifique vue par les sciences sociales

Les sciences sociales apportent un éclairage différent sur l'activité scientifique. Elles ne quantifient pas la reproductibilité des études biomédicales mais s'intéressent à la production des connaissances scientifiques et plus généralement à l'activité scientifique. Même si les sociologues des sciences sont divisés quant à leur définition de l'activité scientifique (Collins 1983; Dubois 2013; Hessels 2008; Pestre 2004; Shinn 2002; Weingart

1997), il n'en reste pas moins que les cadres de références proposés par Merton (Merton 1973) et certains des outils développés par les auteurs de la « nouvelle production des connaissances » (Gibbons et al. 1994) sont utiles pour appréhender nos objets de recherche. En effet, même si en sociologie des sciences l'existence d'un mode 1 ou d'un mode 2 de production de connaissances est vivement critiquée (Jensen et al. 2008; Shinn 2002), compte-tenu de notre objet de recherche, le mode 2 est ici un bon modèle d'étude de l'activité scientifique et plusieurs éléments que nous décrivons au cours de cette introduction et dans la conclusion (Chapitre V) viennent légitimer l'utilisation de ce « nouveau mode de production des connaissances ». De plus, le cadre fourni par ces modes de production de connaissances nous a semblé d'autant plus intéressant que les chercheurs en sciences biomédicales ont tendance à percevoir leur activité comme régie par les normes mertonniennes et comme séparées de la société (Anderson et al. 2007). Enfin l'évolution de la pratique de l'activité scientifique, notamment dans le domaine de la recherche biomédicale, semble répondre à des attributs du mode 2 (voir paragraphe 2.2 de ce chapitre et le chapitre V).

Par conséquent, dans ce paragraphe, nous décrivons les modes 1 et 2 de production des connaissances et explicitons pourquoi nous pensons que ces cadres sont utiles pour décrire l'évolution de l'activité scientifique. Nous présentons également la définition de l'incertitude et la façon dont elle est étudiée en sociologie des sciences. Enfin, nous décrivons l'implication des politiques d'évaluation, du *benchmarking* et de la sociologie de la promesse dans l'activité scientifique et dans l'évolution vers un mode 2 de la production de connaissances.

2.1- Ethos des sciences. Les normes Mertonniennes

Dans l'ouvrage « *the normative structure of science* », Merton décrit l'éthos des sciences et les normes qui s'y appliquent (Merton 1973). Selon Merton, ces normes permettent de codifier socialement le rôle du scientifique, de configurer la science en tant que système social tout en rendant compte de son autonomie. Dans ce cadre normatif, l'activité scientifique a pour seule fonction de produire des connaissances. Pour Merton, les normes et valeurs qui permettent un « bon » fonctionnement de la science sont composées de quatre impératifs institutionnels : l'universalisme, le communisme, le désintéressement et le specticisme organisé (Merton 1973).

- L'*universalisme* : un fait scientifique est accepté ou réfuté indépendamment des attributs sociaux ou personnels de celui qui les propose.

- Le *communisme* : toute découverte scientifique est un bien commun, produit par la collaboration entre scientifiques. Ceci implique donc que toute découverte scientifique doit être publiée et ainsi profiter au plus grand nombre.

- Le *désintéressement* : la production de connaissance est la raison pour laquelle le scientifique réalise ses recherches. Il n'a pas pour objectif principal son propre profit ou sa réussite personnelle. C'est la reconnaissance de son travail par ses pairs qui lui donnera indirectement une bonne réputation et permettra l'avancement de sa carrière.
- Le *scepticisme organisé* : les scientifiques doivent avoir une attitude critique vis à vis de tout énoncé scientifique qui doit être examiné, vérifié, confirmé avant d'être intégré au savoir, à la connaissance. La reproductibilité des résultats est donc une valeur majeure de cet ethos.

Pour Merton, cet éthos représente un idéal de l'activité scientifique d'un point de vue institutionnel, il ne cadre pas forcément avec les motivations des chercheurs. Son caractère normatif est intéressant mais pose problème pour comprendre les comportements déviants. Des contre-normes lui ont rapidement été associées (Mitroff 1976) notamment dans des contextes de forte compétition. Elles permettent de mieux expliquer les motivations et comportements des chercheurs. Ainsi à la norme d'*universalisme* s'adosse la contre-norme de *particularisme* : les résultats peuvent être jugés plus crédibles lorsqu'ils sont publiés dans des journaux prestigieux ou par des équipes de recherche connues. A la norme de *communisme* s'adosse celle de *secret*, les scientifiques protègent leurs résultats pour être les premiers à publier, mais aussi dans le cas des innovations technologiques et de la propriété intellectuelle. A la norme de *désintéressement* s'adosse celle d'*intérêt personnel* : les scientifiques cherchent à publier pour obtenir des financements et favoriser leur carrière. Enfin, à celle du *scepticisme organisé* s'oppose celle du *dogmatisme organisé* : les chercheurs restent attachés à leurs hypothèses, qu'ils défendent souvent au-delà du raisonnable, et doutent des hypothèses des autres chercheurs (Mitroff 1976). En fonction de l'activité de recherche pratiquée par les chercheurs et de son contexte, les normes et contre-normes s'appliquent différemment. C'est ce que Merton avait appelé le « concept de l'ambivalence sociologique » (Bucchi 2015). A la suite d'entretiens approfondis avec des scientifiques, certains auteurs ont montré que ces normes étaient insuffisantes pour décrire leur comportement (Anderson et al. 2010). Ils ont donc ajouté deux autres normes : la norme de *gouvernance* qui implique que les scientifiques sont responsables du contrôle et de l'orientation de la recherche grâce au système de validation par les pairs et la norme de *qualité* : les scientifiques jugent les contributions de leur collègues sur la qualité de celles-ci (Anderson et al. 2010). A ces normes s'opposent bien sûr leur contre-normes, celle d'*administration* où la recherche est gérée par des administrateurs, et celle de *quantité* où les scientifiques jugent leurs collègues sur le nombre de leurs publications et le volume de leurs financements sur appel à projet (Anderson et al. 2010).

Cet ensemble de normes et de contre-normes permet un cadrage des objets de recherche. En effet, si l'on se souvient des pratiques de recherche discutables, on voit en quoi elles dérogent aux normes décrites plus haut et en quoi elles s'inscrivent aisément dans les contre-normes. Par exemple la sélection des seules observations allant dans le sens de l'hypothèse représente une violation flagrante de la norme de scepticisme et relève d'une contre-norme : le dogmatisme. De même les indices bibliométriques utilisés pour l'évaluation des chercheurs relèvent plus de la norme de quantité que de celle de la qualité.

La communauté scientifique se prévaut de son adhérence aux normes décrites par Merton pour garantir le modèle d'autorégulation et d'autocorrection de son activité. Même si les scientifiques ne respectent pas toujours ces normes, celles-ci sont reconnues au sein de la communauté et sont mises en avant pour légitimer l'autonomie et le rôle des sciences (Lam 2010). L'activité scientifique apparaît ici comme une entité indépendante où les scientifiques sont à la fois producteurs, consommateurs et évaluateurs des connaissances scientifiques.

Des interviews réalisées auprès des chercheurs ont d'ailleurs montré qu'ils se reconnaissent dans ces normes et qu'ils essaient de les respecter (Anderson et al. 2007; Anderson et al. 2010; Lam 2010). Mais ils indiquent aussi qu'en fonction de leur environnement de travail et de la perception qu'ils en ont (milieu de recherche coopératif ou compétitif), ils ressentent des difficultés à respecter les normes auxquelles ils sont pourtant attachés (Anderson et al. 2007). Les interviews révèlent aussi que les scientifiques interrogés pensent que les contre-normes s'appliquent plutôt aux comportements de leurs collègues (Anderson et al. 2010). Les auteurs concluent que cette discordance entre l'idée que les chercheurs se font des normes qui régissent leur activité et les comportements qu'ils observent auprès de leurs collègues sont source de stress (Anderson et al. 2007).

Cependant, le modèle normatif de Merton où la science évoluerait indépendamment et sans influence extérieure ne permet pas d'expliquer l'apparition des contre-normes, des pratiques de recherche discutables ou de cette culture de l'évaluation scientifique. Il permet néanmoins de constater que les normes décrites sont tenues en haute estime par les scientifiques et que ceux-ci s'efforcent d'y adhérer. Ces normes permettent d'expliquer la réaction des scientifiques face à la crise de la reproductibilité et leur apparente adhésion aux mesures proposées pour y faire face. Elles montrent aussi que, même si le système de validation par les pairs a des défauts, il fonctionne néanmoins puisque lorsque la norme de scepticisme organisé semble n'être plus respectée, les scientifiques s'en émeuvent. D'autre part, la rhétorique de leur argumentation concernant la crise de la reproductibilité est basée

sur le non respect de ces normes : désintéressement, universalisme, scepticisme organisé, gouvernance et qualité.

2.2- Du mode I au mode II de production des connaissances ?

Selon plusieurs auteurs, la science et la production de connaissances se sont modifiées du fait de la transformation des relations entre les sciences, la politique, les médias et la société (Bonneuil 2004; Bruno 2012; Bucchi 2015; Dubois 2013; Elzinga 1997; Gibbons 1999; Nowotny et al. 2003; Pellizzoni 2003; Wolton 1997).

Ainsi selon Gibbons et al, un nouveau mode de production des connaissances ou mode 2 se serait développé (Gibbons et al. 1994), et coexisterait avec le mode 1, caractérisé par les normes décrites par Merton (Merton 1973).

Les auteurs s'inscrivant dans cette perspective de recherche analysent donc la production et la construction même de l'activité scientifique en accord avec les intérêts sociaux, politiques et économiques des différentes parties prenantes. Dans leur description de ce mode 2, ils distinguent cinq attributs qui caractérisent l'évolution de l'activité scientifique. Premièrement, la recherche est de plus en plus tournée vers l'application, vers la résolution de problèmes. Deuxièmement, la science devient transdisciplinaire, c'est à dire que l'activité de recherche et les groupes qui se forment ne relèvent plus des disciplines. Troisièmement, la construction du savoir devient hétérogène. Elle ne se fait plus au sein des universités mais au sein de réseaux hétérogènes transitoires qui rassemblent des acteurs avec des perspectives différentes (ministères, entreprises, universités, groupes de pression). Le quatrième attribut du mode 2 est l'apparition de nouvelles formes du contrôle de la qualité, l'examen par les pairs n'est plus suffisant. Des critères sociaux, politiques et économiques entrent aussi en jeu. Enfin, la production de connaissances doit être réflexive. Elle ne doit plus être simplement fiable, elle doit être « robuste socialement ». Un nouveau « contrat social » est ainsi défini, les interactions entre science et société ne se font plus dans un seul sens. Elles se font non seulement du scientifique qui produit la connaissance vers la société mais, inversement, la société est désormais en mesure de questionner, voire contester, les scientifiques et en droit d'influencer l'orientation de la recherche (Gibbons 1999; Nowotny 2004). Dans cette agora, les médias ont un rôle de plus en plus important.

Ce modèle de nouvelle production de connaissances a cependant été vivement critiqué par de nombreux auteurs (Bérard et al. 2015; Hessels 2008; Shinn 2002; Weingart 1997). Il n'en reste pas moins que certains des éléments décrits par les promoteurs du mode 2 semblent être au moins partiellement acceptés, comme l'orientation croissante de la recherche vers ses applications potentielles et la notion de responsabilité des scientifiques face à la société et ses différents acteurs (Hessels 2008). D'autre part, malgré les critiques émanant de certains

sociologues des sciences, cette description d'un nouveau mode de production des connaissances « fournit un langage efficace pour orienter l'action publique » (Dubois 2013).

Nowotny et Gibbons ont expliqué l'apparition de ce nouveau mode de production du savoir par la coévolution de la science et de la société (Gibbons 1999; Nowotny 2004). Pour Nowotny, cette coévolution est loin d'être synonyme d'harmonie. Pour cette sociologue, elle se fait dans le contexte d'une économie de marché où « le niveau sans précédent d'éducation de nos sociétés, l'omniprésence des technologies d'information et de communication, la réalisation que la production d'incertitudes est inhérente à ce processus de coévolution font que la société est dans une position idéale pour exprimer ses souhaits, ses désirs et ses peurs à la Science » (Nowotny 2004).

John Ziman va plus loin et parle de science « post académique ». Pour lui, ce n'est pas seulement un nouveau système de production des connaissances mais une nouvelle « culture » de la recherche (Ziman 1996). La croissance de l'activité de la recherche est telle que les financements ne peuvent plus être assurés intégralement par les gouvernements. Par conséquent, les scientifiques se doivent d'être efficaces et de « vendre » leurs résultats. Leur rôle est de produire des connaissances utiles, pratiques, applicables. Leur activité est donc moins orientée vers la production de connaissances que vers la résolution de problèmes. Cette science post-académique s'inscrit d'autre part dans un système de communication où l'information est transmise rapidement, que ce soit entre les groupes de recherche, les institutions gouvernementales ou le grand public. La recherche doit être finalisée, entraînant la mise en place de programmes de recherche limités dans le temps avec des concepts et techniques établis. Dès lors, la norme de désintéressement est impossible à soutenir et, avec les financements privés, les conflits d'intérêts émergent. Comme pour le mode 2, l'activité scientifique est produite, construite et répond désormais à une culture où le pouvoir socio-économique est plus important que les valeurs scientifiques.

Ziman soulève aussi un point important. Il se demande si cette « science post-académique » ne va pas perdre son objectivité (Ziman 1996b). Pour lui, cette objectivité scientifique reste une valeur essentielle en tant que garante de la qualité et de l'impartialité des connaissances produites. Même si il convient que l'objectivité totale ne peut être atteinte, il insiste sur le fait que la recherche post-académique est orientée d'abord vers la résolution de problèmes, l'efficacité et la profitabilité et qu'il semble alors difficile de maintenir, dans de telles conditions, l'impératif de reproductibilité.

En somme, quelque soit le nom qu'on lui donne, la science du « mode 2 », « post-académique » ou encore « la science 2.0 » (Bucchi 2015) est modifiée en raison de l'incorporation de facteurs comme les intérêts commerciaux, les politiques d'évaluation ou

l'intérêt sociétal à la pratique scientifique. La recherche serait de plus en plus menée non pas pour poursuivre la production de connaissances mais pour « répondre » à des demandes économiques ou sociétales.

Cette nouvelle production des connaissances serait également associée à une plus grande orientation vers les médias. La science du mode 1 « boudait » les médias et les chercheurs des années 1950 et 1960 estimaient que les interactions avec la presse auraient un impact négatif sur leur carrière ou sur l'opinion de la communauté scientifique à leur égard (Boltanski et al. 1970). Au contraire, la science du mode 2 considère les médias comme des interlocuteurs importants. Peut-être en raison de l'influence donnée à la visibilité médiatique dans l'orientation des décideurs politiques ou des investisseurs financiers (Bucchi 2015).

Nous pensons donc que la description des modes 1 et 2 ou 2.0 de production des connaissances fournit de bons éléments de cadrage pour expliquer l'évolution de l'activité scientifique. Notamment, nous voyons que dans cette « nouvelle production des connaissances », la reproductibilité des résultats n'est plus centrale à l'activité scientifique et que des intérêts externes (évaluation et financement de la recherche, visibilité médiatique) entrent en jeu.

2.3- Reproductibilité et incertitude : éclairages de la sociologie des sciences

Comme nous l'avons déjà décrit en utilisant pour référence l'ouvrage *Agir dans un monde incertain*, l'incertitude en sociologie des sciences est celle liée aux sciences émergentes pour lesquelles les conséquences ne sont pas connues ou les connaissances manquent. Autrement dit, l'incertitude ici est celle liée aux controverses (Callon et al. 2014). C'est pourquoi, pour les sociologues des sciences, l'étude des controverses constitue un outil de choix pour comprendre la production de faits scientifiques (Collins 1983; Collins 1987). Ces controverses peuvent être d'ordre purement scientifique : il y a un désaccord entre les chercheurs (par exemple la mesure des ondes gravitationnelles ou des neutrinos solaires) ou alors la controverse touche la sphère publique et les résultats scientifiques ne peuvent suffire pour légitimer la décision (Collins et al. 2002; Pinch 1981). Dans le cas de l'incertitude liée à la prise de décision, les auteurs questionnent la légitimité des sciences à diriger et orienter la décision (Collins et al. 2002; Pellizzoni 2003) et parlent de co-construction des savoirs (Jasanoff 1996). Ils montrent par exemple que dans le cas de l'encéphalite spongiforme bovine en Angleterre (Irwin 2006), des conséquences de Chernobyl (Wynne 1996) ou encore du SIDA (Epstein 1995), le savoir et l'expertise scientifique ne sont pas suffisants pour que la décision soit légitime : l'opinion de la société, au sens plus ou moins élargi, ne doit pas être négligée. Un résultat scientifique ne devient donc pas un fait légitime sans co-construction avec des acteurs non-scientifiques. On voit ici deux choses : premièrement, ces études

concernent des résultats de recherche controversée. Elles ne concernent pas le type de recherche biomédicale que nous voulons analyser ici. Deuxièmement, l'incertitude étudiée n'est pas celle qui nous intéresse. Elle ne tient pas compte de l'incertitude liée aux résultats initiaux et donc de la réplication indispensable d'un résultat pour en assurer la validité.

Pourtant, Gibbons mentionne l'importance de la réplication et donc indirectement l'incertitude lorsqu'il utilise la définition de Ziman d'un fait scientifique : un fait scientifique est robuste lorsque ses résultats ont été répliqués et qu'un consensus a été atteint entre les pairs (Gibbons 1999). D'autre part Collins et al, aussi opèrent une distinction : il y a la « science normale non controversée » pour laquelle la méthode scientifique classique de validation et de vérification des résultats par les pairs est suffisante, pour laquelle il n'y a pas de disputes majeures (Collins et al. 2002) et il y a la science controversée pour laquelle le simple fait de reproduire les expériences ne permet pas d'expliquer ni de résoudre les différences d'opinion entre scientifiques (Collins 1983). Enfin, pour Godin et Gingras, le consensus scientifique est obtenu lorsqu'il n'y a plus d'objections majeures à éliminer et qu'aucune explication plus adéquate ne peut être avancée (Godin et al. 2002). Ce consensus se définit donc devant une audience qui suit des règles établies pour critiquer ou valider les observations. La communication des résultats à cette audience est indispensable à l'évaluation. Le fait pour la communauté scientifique d'essayer de reproduire l'expérience est primordial pour pouvoir apporter de nouvelles connaissances : la réplication permet de confirmer le résultat ou d'avancer de nouvelles hypothèses ou conclusions.

Les sociologues des sciences semblent donc être familiers avec l'importance de la réplication et la définition d'un fait scientifique. Pour autant, ils ne s'intéressent pas à la « science normale non controversée » et aux conséquences de la non reproductibilité des sciences sur le public et les décideurs. Lorsqu'ils s'intéressent à la reproductibilité des sciences c'est pour introduire la théorie de la régression de l'expérimentateur. Ainsi, en utilisant l'exemple de la controverse scientifique qui entoure la découverte des ondes gravitationnelles, Collins explique que si l'expérience de Weber ne pouvait pas être reproduite c'est parce que la définition d'un bon outil de mesure manquait et que les connaissances tacites des scientifiques opéraient dans ce contexte. La théorie de la régression de l'expérimentateur permet d'expliquer que les mesures ne puissent être reproduites : si l'outil de mesure n'est déclaré efficace que quand il permet de mesurer le résultat attendu et qu'on utilise cet outil pour vérifier le résultat alors la résolution est impossible (Collins et al. 2001). Cependant, si nous reprenons l'exemple du polymorphisme génétique associé à une pathologie dans une certaine population, l'outil de mesure ne peut pas être impliqué. Le séquençage génétique est une technique largement utilisée par de nombreux chercheurs.

Ainsi pour ces sociologues, la définition d'un fait scientifique pour la science « normale » ne pose pas problème : le consensus est atteint après répllication et validation des résultats par les pairs. Selon eux, l'incertitude qui mérite d'être étudiée est celle de la science controversée, soit parce que le consensus n'est pas atteint du fait d'un désaccord persistant entre scientifiques, soit en raison de résultats scientifiques jugés comme insuffisamment légitimes pour justifier une décision publique. Pourtant, nous le verrons au cours de ce manuscrit, il est primordial de tenir compte et d'étudier la présentation de l'incertitude liée aux résultats initiaux puisque ce sont bien souvent eux qui sont les plus visibles, tant pour la communauté scientifique que pour les médias médiatique et donc pour le public et les décideurs politiques.

2.4- Les politiques de recherche et l'évaluation de la recherche

D'autres aspects nous semblent importants pour indiquer l'évolution de l'activité scientifique vers un mode 2 de production des connaissances : les politiques de recherche et d'évaluation basée sur des indices numériques. En effet, en France, la SNR, Stratégie Nationale de Recherche France-Europe 2020 a proposé dix défis sociétaux « pour répondre au défi permanent de la connaissance et aux défis socio-économiques inédits » (Ministère de l'Education Nationale 2015). Parmi ces défis, le défi numéro 3, *Santé et Bien-être* nous semble tout à fait pertinent à relever dans le cadre de cette thèse. L'accent est mis en particulier sur les « déterminants environnementaux, sociaux et économiques des maladies et de leur retentissement sur la personne et sur la société. (...) La recherche en santé est naturellement un enjeu majeur de politique publique, et un vecteur déterminant de développement économique ». On retrouve ici certains des traits exposés dans la stratégie de Lisbonne critiqués par Isabelle Bruno. Dans sa critique de cette stratégie, l'activité scientifique n'est plus synonyme de production de connaissances mais de production de connaissances utiles ou exploitables économiquement (Bruno 2012). Il faut créer « le marché commun de la recherche ». Pour cela, non seulement les connaissances produites doivent être applicables mais les chercheurs doivent pouvoir « circuler librement » et devenir des « chercheurs européens ». Le chercheur est alors placé en situation de concurrence et devient un « chercheur-entrepreneur » qui doit avoir « une utilité sociale ». Il est évalué selon des indicateurs de performance. Ces mesures statistiques de l'activité des chercheurs découlent directement des nouveaux outils d'évaluation de la productivité inspirés du *benchmarking*⁸.

⁸ Benchmarking « qui veut s'améliorer doit se mesurer, qui veut être le meilleur doit se comparer ». Cité dans « A vos marques®, prêts cherchez ! » (Bruno-2012)

Isabelle Bruno en critique justement les dérives avec sa formule: « ne cherchez plus, innovez ! » (Bruno 2013). Pour elle, « accorder la primauté à l'innovation comme finalité ultime de la recherche, c'est ne pas avoir saisi la nuance qui existe entre le fait d'être utile et l'intention qui pousse à l'être. L'utilité du chercheur en société réside dans l'exercice même de son métier, lequel consiste à construire des savoirs » (Bruno 2013).

Bruno donne ainsi un aperçu de l'influence des politiques de recherche sur l'activité des chercheurs. Elle montre que dans cette « économie de la connaissance » les conditions de travail des chercheurs ont été bouleversées du fait des nouvelles temporalités de la recherche, du financement sur projet et de la compétition qui en découle. Ainsi, une enquête a été réalisée sur l'opinion des scientifiques d'un laboratoire de physique dans le domaine des nanosciences sur le financement « sur projet » (Jouvenet 2011). Elle montre que les scientifiques, dans une situation où les coûts de recherche ont augmenté et sont associés à des baisses de dotations récurrentes, se trouvent dans l'obligation de faire de la « pêche aux contrats ». Ils révèlent aussi qu'ils « suivent des projets à la mode » même si ça « ne cadre pas avec leur objectifs », que les contrats sont orientés vers la recherche appliquée et qu'ils cadrent mal avec « leur liberté ». Ils se plaignent de l'extension des activités administratives et managériales associées au financement sur contrat et de la perte « macroscopique de temps associée » à rédiger des « projets famélicux » puis à être « rapporteurs pour les dossiers qui ont été faits par d'autres ». Ils soulignent aussi l'importance de bien communiquer leurs travaux de recherche pour « séduire des financeurs » quitte à « vendre des résultats déjà obtenus » ou « survendre leur projet en le fondant sur des promesses aussi belles qu'irréalistes » (discuté plus loin avec la sociologie de la promesse). On reconnaît ici un des attributs du mode 2 : l'orientation de la recherche vers les applications. Les contre-normes d'administration et d'intérêt personnel interviennent alors dans ce contexte.

Par ailleurs, l'évaluation des politiques publiques de recherche peut se faire en partie sur la base de l'évaluation de la production des connaissances scientifiques et donc de ceux qui la produisent, les chercheurs. Ainsi les indicateurs bibliométriques sont des outils très fréquemment utilisés (Fouquet et al. 2010). Ces indicateurs quantitatifs, malgré leur facilité d'emploi, sont souvent critiqués en raison des effets pervers qu'ils induisent (Osterloh et al. 2008). En effet, comme nous l'avons vu plus haut, les chercheurs sont évalués sur leur production d'articles scientifiques, sur la qualité de ces publications (la qualité étant ici le facteur d'impact du journal dans lequel les articles sont publiés), sur leur indice h ⁹ et sur le

⁹ L'indice h ou indice de Hirsh est une mesure de la productivité et de l'impact des recherches du chercheur. Il prend en compte le nombre d'articles écrits par les chercheurs et le nombre de fois que chacun des articles a été cité.

nombre de financements qu'ils ont obtenus. Cette pratique d'évaluation a donné naissance à l'expression « *Publish or Perish* » ou course à la publication. Ainsi, dans une étude scientométrique analysant les recrutements de scientifiques au CNRS entre 2005 et 2013, Brischoux et Angelier ont montré que, non seulement les personnes recrutées actuellement avaient une plus grande expérience en recherche que leurs prédécesseurs, mais aussi qu'elles avaient un nombre plus élevé de publications (Brischoux et al. 2015). Merton avait déjà parlé de « l'effet Saint Matthieu » : les chercheurs publiant dans les journaux scientifiques prestigieux ont accès plus facilement aux financements et donc en retour aux publications dans ces mêmes journaux. Les contributions de scientifiques éminents auraient donc une influence importante sur l'orientation des activités de recherche en permettant d'attirer l'attention du reste de la communauté scientifique sur ces sujets. En effet, non seulement la publication dans les journaux prestigieux garantit une reconnaissance presque indiscutable aux yeux des autres scientifiques, mais elle donne une visibilité médiatique (Bucchi 2015). Dès lors se dessine l'importance de publier des résultats extraordinaires (pour susciter l'attention des médias) dans des journaux prestigieux (valeur plus importante pour l'évaluation du chercheur) afin d'orienter indirectement l'activité scientifique et les politiques de recherche¹⁰.

2.5- Sociologie de la promesse

Un dernier courant de recherche en sociologie vient, à notre avis, expliquer en partie la crise de la reproductibilité observée en sciences : la sociologie de la promesse. Comme l'enquête de Jouvenet le souligne, les chercheurs ont modifié leur comportement face à l'évolution des politiques de recherche. Ils ont désormais tendance à « survendre » leurs projets et à promettre des résultats extraordinaires (Jouvenet 2011). Ces comportements peuvent être éclairés par la sociologie de la promesse. En effet, si l'on prend l'exemple de la recherche biomédicale qui nous intéresse ici, elle est évidemment porteuse de promesses. Elle est orientée vers la compréhension et le traitement des symptômes des maladies. Elle est donc susceptible de produire de nouvelles opportunités pour de nombreux acteurs (patients, médecins, industrie pharmaceutique, associations de patients, assurance maladie etc.). Mais elle doit être financée, et c'est à ce niveau qu'interviennent les promesses. Les analystes de l'innovation expliquent leur importance : elles sont des abstractions orientées vers le futur. Les chercheurs ne savent pas ce qu'ils vont trouver ni ce qu'ils vont produire. Ils ont des « visions », des « attentes » (en anglais *expectations*). La formulation et la présentation de ces attentes sont importantes pour structurer les activités de recherche, légitimer les projets, attirer l'intérêt et sécuriser les financements (Rusconi et al. 2014). Ces attentes, selon Borup, sont le

¹⁰ Ce point sera plus amplement discuté dans le chapitre V.

« chaînon manquant » qui permet de faire correspondre une question technique à une question sociale et de faire interagir différents groupes d'acteurs (Borup et al. 2006). Ainsi, pour obtenir des financements, les scientifiques doivent « vendre » leur recherche, promettre des résultats. La frontière entre la promesse de résultats raisonnables (*hope*) et la survente de résultats (*hype*) est alors mince. Une analyse de Brown montre que dans le cas des biotechnologies, les bénéfices sont souvent exagérés et les risques minimisés (Brown 2003). Ainsi, la promesse de résultats incertains tend souvent à l'exagération, utilisée pour susciter l'intérêt. Ce processus est décrit par le modèle du *hype* de Gartner (Rusconi et al. 2014; van Lente et al. 2013) : après la phase de promesses et de présentations des attentes exagérées (*cycle de hype*), vient une phase de « désillusion » lorsque les promesses ne sont pas atteintes. Les attentes sont alors réajustées en fonction des résultats déjà obtenus. On voit alors que ce qui est crucial c'est la façon dont ces attentes sont communiquées et présentées à la société. Ainsi une étude sur les promesses de la génétique « *genohype* » (Caulfield 2004) montre que la survente des résultats entraîne une perte de confiance de la société et des investisseurs. Certaines communautés scientifiques ont pris conscience des risques induits par des promesses exagérée. Ainsi, la société internationale de la recherche sur les cellules souches (ISSCR) a insisté sur l'attention que les scientifiques doivent porter à leur discours public sur ces sujets (Caulfield et al. 2016). Elle encourage les scientifiques à ne pas extrapoler des résultats préliminaires en parlant de futurs essais cliniques. Une analyse des discours des neuroscientifiques sur l'imagerie cérébrale notamment a aussi conduit à recommander aux scientifiques de tempérer la présentation de leurs attentes en gardant à l'esprit l'impact potentiel de leurs « visions » sur les attentes de la société (Rusconi et al. 2014).

D'autres études se sont intéressées plus directement aux conséquences de la survente, de l'exagération de la portée ou de l'intérêt sociétal des recherches biomédicales sur la confiance du public et donc sur la recherche elle-même (Master et al. 2013). Si la survente est une étape fréquente du développement technologique et que les scientifiques ont intérêt à y avoir recourt lors de leur communication avec le public, les auteurs soulignent qu'une connotation péjorative lui est aussi associée. En effet, cette survente peut aussi qualifier des comportements contraires à l'éthique. Selon les auteurs, quatre problèmes sont associés à la survente. Pour le public, la survente dans les médias nuit à la compréhension et à la confiance qu'il porte aux scientifiques (1). Au niveau du financement et de la pratique de la recherche, la survente conduit à ce que des scientifiques et des politiques concentrent des ressources autour d'objectifs irréalisables au détriment d'autres projets scientifiques (2). Cette survente peut également entraîner une traduction trop rapide d'une découverte scientifique en applications pratiques mal maîtrisées (Moreira et al. 2005) avec les conséquences néfastes qui

en résultent (3). Enfin, la survente intentionnelle et systématique des résultats représente une brèche dans l'éthique de la recherche (4). Elle produit au mieux des inexacitudes et au pire des résultats faux et non reproductibles.

On voit donc bien que, si la promesse de résultats est importante pour obtenir des financements, elle ne doit pas prendre le pas sur la qualité de la recherche et des résultats obtenus sous peine de remettre en cause l'éthique des scientifiques et donc le financement de la recherche.

En somme, la caractérisation de l'activité scientifique présentée dans ce paragraphe permet de constituer des cadrages pour l'objet de notre recherche. Les scientifiques interrogés déclarent leur attachement aux valeurs des normes décrites par Merton. Mais le contexte grandissant de compétition pour des ressources réduites les conduit à délaissier certaines de ces normes au profit des contre-normes. En effet, les politiques de recherche ont permis de déterminer des grands axes de recherche avec des financements associés. Ces axes constituent des « défis sociétaux » : la recherche est alors tournée vers la résolution de problèmes (réchauffement climatique) ou la production d'innovations (en matière de santé et d'énergie). Les chercheurs doivent donc répondre à des appels d'offres pour pouvoir réaliser leurs recherches. Pour obtenir ces financements, ils doivent présenter un projet compétitif avec un impératif de finalisation et démontrer leur capacité à mener à bien ce projet. L'obtention des financements est donc corrélée à la qualité du sujet, jugée en fonction de l'impact des résultats attendus sur la société, mais aussi à la « qualité » des chercheurs, mesurée en terme d'indices bibliométriques. Les inconvénients de ces politiques de recherche sont importants : les sujets abordés sont des sujets « peu risqués » permettant d'obtenir rapidement des publications, les sujets plus créatifs sont donc délaissés. D'autre part, les incitations à la réplique des résultats sont inexistantes : la réplique d'une étude ne permet pas de publication dans des bonnes revues et sans publication, il n'y a pas de financement. Enfin, la course des scientifiques pour publier dans des journaux prestigieux est effrénée : publier dans *Science* ou *Nature* assure l'attention des autres équipes de recherche, des agences de financements, des médias et donc des décideurs politiques. Tout un tas de pratiques discutables entrent alors en jeu comme la « vente de promesses » irréalisables et déconnectées du sujet pour pouvoir obtenir un financement. Les conséquences sociétales peuvent être importantes, la confiance du public peut être ébranlée.

D'autre part, l'incertitude liée aux résultats initiaux n'a pas encore été étudiée. L'incertitude n'est considérée ici qu'en terme de controverses, de désaccord entre les

scientifiques ou d'impact sur la société. Dans la science dite « normale », l'exigence de reproductibilité garantit la fiabilité des résultats. Il n'est pas envisagé que des résultats publiés dans des grands journaux puissent être exagérés, ni trop « visibles » pour la communauté scientifique et les médias, donnant ainsi une crédibilité trop importante à certains projets. Seule la fraude est discutée (Haran et al. 2009).

On constate aussi une évolution croissante du « contrat social » (Gibbons et al. 1994) dans la légitimation des sciences : la science ne doit plus seulement être communiquée au public, elle doit être co-construite. Chacun peut ainsi participer à la démocratie technique et discuter des politiques et orientations de recherche en matière de génétique, de technologie et d'environnement par exemple. Ainsi, Michel Callon a présenté 3 modèles de diffusion des connaissances scientifiques : le modèle de *l'instruction publique* ou d'éducation publique où seules les sciences sont aptes à permettre la décision, le modèle du *débat public* où la place de certains acteurs « profanes » est reconnue pour compléter et enrichir les savoirs scientifiques et enfin le modèle de la *co-production des savoirs* où scientifiques et non-scientifiques interagissent librement (Callon 1998). Ces trois modes de diffusion coexistent et leur utilisation dépend du sujet à présenter. Nous verrons, dans le paragraphe suivant, que ces modes de diffusion sont similaires à ceux décrits par les chercheurs en sciences de la communication.

III- Sciences et Médias

Nous n'avons pas encore évoqué un des acteurs majeurs de la diffusion des connaissances : les médias. Or, ils constituent le lien entre les découvertes scientifiques et le public puisque c'est au travers des canaux journalistiques traditionnels que celui-ci obtient des informations sur les avancées et découvertes scientifiques (Eurobaromètre 2007; National Academies of Sciences 2017). La compréhension des résultats scientifiques et de leurs implications lui vient donc majoritairement de ce qu'il lit dans les médias (Nelkin 1995). La façon dont les sciences sont présentées, les cadres et images choisis par les journalistes ont donc une importance majeure (de Semir 2010).

Plusieurs séries d'études ont été réalisées pour essayer d'appréhender la façon dont les découvertes scientifiques sont sélectionnées et présentées dans les médias. En effet, développer des connaissances sur ces sujets est important dans la mesure où la science fait partie intégrante de la société et intervient à tous les niveaux. Selon Durant et al, il est important de s'inquiéter de ce que le public comprend de la science parce que

« premièrement, la science est une des réussites majeures de notre culture, deuxièmement, la science affecte la vie de tout à chacun, troisièmement, de nombreuses décisions publiques sont faites sur des bases scientifiques et elles ne peuvent être démocratiques que lorsqu'elles sont prises dans un contexte de débat informé. Et quatrièmement, la science est financée par des fonds publics et donc le public doit avoir connaissance des résultats » (Durant et al. 1989).

De nombreuses études ont été réalisées pour mesurer l'intérêt du public pour la science et essayer de le corréler à son niveau de connaissances (Durant et al. 1989; Ziman 1991). La théorie souvent avancée est celle d'un déficit : si le public n'a pas de connaissances scientifiques comment pourrait-il prendre part à la décision ? Le corollaire est que si le public avait une meilleure connaissance et compréhension des sciences, il serait alors capable de prendre des décisions, en accord avec la méthode scientifique (de Semir 2010; Priest 2001).

Dans cette perspective, des auteurs ont critiqué le « centrisme » de l'étude de la communication des sciences sur la promotion de l'agenda scientifique (Dornan 1990; Hansen 2016). Si les études basées uniquement sur la mise en évidence de différences entre les publications scientifiques et les articles de presse sont intéressantes, elles ont aussi leurs limites. En effet, elles sont basées sur l'idée que se fait le scientifique de la « bonne vulgarisation ». Comme le suggère Hilgartner, il est normal que des articles de *Cell* et du *Baltimore Sun* soient différents. Ce qu'il faut c'est que la « popularisation » du résultat scientifique soit adéquate. Il ne faut pas mesurer des différences selon des critères scientifiques, mais savoir si les simplifications opérées sont acceptables pour la compréhension du public (Hilgartner 1990). Des études ont d'ailleurs montré que les scientifiques acceptaient certains des impératifs journalistiques et donc des inexactitudes ou raccourcis dans la présentation de leurs résultats dans la presse (Gascoigne et al. 1997; Peters 2012b). De nouveaux modèles d'études de la communication des sciences ont été proposés, tenant compte de l'avis public dans la communication et la production des sciences (National Academies of Sciences 2017). Ces modèles ont d'abord parlé de « dialogue » plutôt que de « compréhension » (*public understanding*) puis de « participation du public » (*public engagement*) (de Semir 2010; Schafer 2009). Certains auteurs indiquent que ces différents modes de diffusion des connaissances - compréhension, dialogue et participation - coexistent et dépendent des connaissances que l'on veut diffuser (Trench 2008). On voit aussi que ces modèles décrits dans les sciences de la communication correspondent à ceux décrits par les sociologues : *public understanding* et instruction publique, dialogue et débat public et enfin *public engagement* et démocratie technique.

Par conséquent, afin d'analyser la communication des sciences, plusieurs types d'études ont été réalisés, dans différentes disciplines, notamment en sciences de la communication et en sciences biomédicales. Ces études comparent, par exemple, le contenu des publications scientifiques ou des communiqués de presse avec les articles de presse, ou encore la présentation de l'incertitude et des risques associés (Gonon et al. 2011; Sumner et al. 2016). D'autres ont examiné les facteurs qui influencent le choix des sujets couverts (Badenschier et al. 2012; de Semir 1996; van Trigt et al. 1994). D'autres encore ont analysé les interactions entre les chercheurs et les journalistes et tenté d'expliquer les tensions qui existent entre ces deux catégories d'acteurs (Nelkin 1996; Watts 2014). Enfin, des études se sont concentrées sur les audiences et les modèles de communication des résultats scientifiques (Hilgartner 1990; Trench 2008).

Nous présentons ici les conclusions de certaines de ces études sur la qualité de la présentation des sciences biomédicales, aussi bien dans la presse grand public que dans les communiqués de presse - qu'ils soient issus des institutions ou des revues scientifiques – et que dans les publications scientifiques elles-mêmes. Nous décrivons l'évolution temporelle des interactions entre les scientifiques et les journalistes. Nous introduisons aussi le concept de *médialisation* qui permet d'expliquer en partie l'évolution positive de ces interactions et nous finissons par évoquer la définition de l'incertitude pour les auteurs en sciences de la communication.

3.1- Couverture médiatique des sciences

De nombreux articles s'interrogent sur la qualité de la présentation des sciences dans les médias (Brown 2012; Dentzer 2009; Russell 1999). En effet, les scientifiques ont tendance à reprocher aux journalistes des inexactitudes et des exagérations lorsque ceux-ci couvrent les sciences (Condit 2004; Hilgartner 1990).

Ainsi des études se sont intéressées à la couverture médiatique de différents sujets comme les facteurs de risques génétiques (Holtzman et al. 2005), la génomique (Geller et al. 2005; Kua et al. 2004; Rodder 2009), les traitements (Ankney et al. 1996; Moynihan et al. 2000), les essais cliniques (Yavchitz et al. 2012), les neurosciences (Gonon et al. 2011). La plupart de ces études ont comparé les contenus des publications scientifiques avec ceux des communiqués de presse et des articles de presse. Ces études ont montré que des différences apparaissaient : dans les articles de presse, il manque souvent des informations sur les coûts des traitements, les conflits d'intérêt ou les financements des recherches (Holtzman et al. 2005; Moynihan et al. 2000; Schwitzer 2008), sur les méthodes (Ankney et al. 1996; Pellechia 1997) ou encore le contexte (Friedman et al. 1999). Dans le cas des essais cliniques, le

bénéfice des traitements est souvent exagéré alors que les effets secondaires indésirables sont soit omis, soit minimisés (Brechman et al. 2009; Iaboli et al. 2010; Yavchitz et al. 2012).

Dans l'ensemble, ces études montrent qu'il existe de nombreuses exagérations et distorsions, tant au niveau des titres des articles de presse qu'au niveau de leur contenu (Bubela et al. 2009; Woloshin et al. 2009). Cependant, ces exagérations ne sont pas toujours le fait des journalistes. En effet, ces exagérations et distorsions sont la plupart du temps déjà présentes dans les communiqués de presse des institutions de recherche (Sumner et al. 2014; Woloshin et al. 2009 b), des journaux scientifiques (Sumner et al. 2016; Woloshin et al. 2002; Yavchitz et al. 2012) et même dans les publications scientifiques (Gonon et al. 2011; Woloshin et al. 2009). *A contrario*, si le communiqué de presse n'est pas exagéré, l'article de presse ne le sera pas non plus (Schwartz et al. 2012). Cette exagération dans les communiqués de presse est d'autant plus injustifiée que, contrairement à une idée reçue, elle ne semble pas influencer l'intensité de la couverture médiatique (Sumner et al. 2016). Elle est cependant problématique parce que les communiqués de presse sont une source privilégiée pour les journalistes (Bartlett et al. 2002; de Semir et al. 1998; Stryker 2002). Les institutions scientifiques et les chercheurs sont bien souvent les principaux responsables des exagérations et distorsions présentes dans les communiqués de presse puisqu'ils sont rédigés au sein de ces institutions et, le plus souvent, sous le contrôle des auteurs de la recherche.

Enfin, l'exagération des résultats de la recherche se fait aussi parfois à un autre niveau. En effet, lorsque les chercheurs sont interviewés par des journalistes sur leur recherche, ils ont tendance à la promouvoir en extrapolant leurs résultats ou en parlant d'implications thérapeutiques (Holtzman et al. 2005; Jensen 2010; Nelkin 1996; Woloshin et al. 2009). Par exemple, en parlant de la mise sur le marché d'un traitement contre le cancer dans les deux ans alors que les essais thérapeutiques n'ont été réalisés que chez la souris ou encore en décrivant la découverte d'un polymorphisme génétique impliqué dans une maladie comme un test diagnostique fiable ou une perspective de traitement.

Un autre problème majeur de la couverture des sciences, mais qui a été peu étudié (Gonon et al. 2012), est que les journalistes, en utilisant comme source les communiqués de presse ou les résumés rédigés par les grands journaux scientifiques (Entwistle 1995; Suleski et al. 2010) favorisent la couverture médiatique de résultats préliminaires. En effet, les grands journaux scientifiques ont tendance à publier des résultats nouveaux et attractifs pour la communauté scientifique et les médias (Bucchi 2015). Ces résultats sont donc souvent des résultats initiaux qui ne sont pas encore confirmés par des études ultérieures. Ainsi un conseil récurrent des chercheurs s'intéressant à la médiatisation des sciences biomédicales est de

mentionner, dans les articles de presse, l'importance de répliquer les résultats (Bubela et al. 2009; Holtzman et al. 2005).

Au total, ces études indiquent que la couverture médiatique des recherches biomédicales est beaucoup plus intense que pour les autres domaines scientifiques. En effet, les journalistes couvrent fréquemment des études scientifiques en rapport avec la santé que ce soit dans le cas des traitements, des facteurs de risque ou des modes de vie (Hansen 1994; Pellechia 1997; Stryker 2002; Suleski et al. 2010). Cependant, les scientifiques impliqués dans les recherches présentées considèrent souvent l'information donnée comme étant « de mauvaise qualité ». Ainsi, pour améliorer la qualité de la couverture médiatique des recommandations ont été faites. Les journalistes devraient être plus critiques vis à vis de leurs sources (de Semir 2010; Dentzer 2009), et plus conscients des « valeurs et des vices de la science » (Watts 2014). Les scientifiques devraient aider les journalistes dans leur appréciation de la qualité de la recherche, indiquer des études contradictoires et parler de réplication (Bubela et al. 2009; Dentzer 2009). Des formations sur le fonctionnement de la recherche et l'analyse des résultats devraient ainsi être proposées aux journalistes, et pour les scientifiques, des formations leur permettant de mieux communiquer leurs résultats (Ankney et al. 1996; de Semir 2010; Holtzman et al. 2005).

3.2- Les relations chercheurs-journalistes

Les relations entre les scientifiques et les journalistes ont largement été étudiées. Elles mettent en lumière des normes et perspectives conflictuelles (Nelkin 1996; Valenti 1999). Selon les normes de Merton, les scientifiques sont supposés accorder une importance particulière à la production de connaissances, à l'obtention de données fiables, quantitatives et reproductibles. Ceci implique un processus long. Les journalistes eux attachent une importance particulière à la nouveauté, au caractère sensationnel ou conflictuel de la recherche. L'important pour eux est d'être lus pas d'être exacts (Laurance 1998). Ils travaillent avec des contraintes de temps importantes.

Les interactions entre les scientifiques et les journalistes ont longtemps été considérées comme problématiques du fait de cette disparité entre leurs normes et beaucoup de recherches ont été réalisées dans le but de « combler ce fossé » (Hartz et al. 1997; Peters 1995; Singer 1990). Ainsi, les scientifiques accusent souvent les journalistes de sensationnalisme (Condit 2004; Cooper et al. 2002). Les journalistes eux ne considèrent pas que leur rôle est de chanter les « louanges de la science » (Nelkin 1996) ni d'être de simples relais (Maille et al. 2010; Peters 1995). Mais si les relations entre les chercheurs et les journalistes ont été difficiles (Peters 2012), elles se sont améliorées ces dernières années et sont jugées comme plutôt

satisfaisantes par les deux parties (Geller et al. 2005). Plusieurs enquêtes auprès des scientifiques de différentes disciplines et de différents pays ont en effet révélé que les scientifiques étaient en général contents de leurs interactions avec les journalistes (Allgaier et al. 2013; Besley et al. 2013; de Semir 2010; Gascoigne et al. 1997; Maille et al. 2010).

Par ailleurs, scientifiques et journalistes s'accordent sur l'importance de ne pas exagérer les résultats et d'être les plus exacts possibles (Schwartz et al. 2004). Les scientifiques pensent d'ailleurs qu'ils ont une part de responsabilité (Geller et al. 2005). Les journalistes eux déclarent qu'ils n'approuvent pas l'exagération et n'aiment pas y recourir (Entwistle 1995) mais que les pressions éditoriales les y obligent souvent (Cooper et al. 2002; Jensen 2010).

Mais alors comment expliquer la couverture sensationnaliste des découvertes scientifiques ? En effet, on peut citer ici quelques titres d'article de presse qui n'ont sans doute pas manqué d'attirer l'œil du lecteur : « un seul verre d'alcool augmente le risque de cancer » *Le Figaro*, février 2009, « Sur la piste du vaccin anti-tabac » *Le Parisien*, juin 2012, « Sclérose en plaques : l'espoir d'un nouveau traitement », *Paris Match*, septembre 2016 ou encore « ce que nous mangeons au petit déjeuner influence notre comportement », *Sciences et vie*, juin 2017. Les chapitres III et IV de ce manuscrit apporteront des éléments de réponse.

3.3- Le concept de *médialisation*

Nous l'avons vu, plusieurs chercheurs en sciences sociales décrivent une évolution de l'activité scientifique (Bucchi 2015; Elzinga 1997; Gibbons et al. 1994; Hessels 2008). L'approche que nous avons plus particulièrement retenue est celle de Gibbons et al, et du mode 2 de production des connaissances. En sciences de la communication, l'allemand Peter Weingart a expliqué cette évolution en utilisant le concept de *médialisation* (Weingart 1997). Selon cet auteur, la *médialisation* est la capacité accordée aux médias de façonner l'opinion, de mettre en lumière et de cadrer les problèmes publics. Ainsi, les scientifiques et leurs institutions considèrent de plus en plus les médias comme des éléments clés pour porter leurs projets à l'attention des politiques et du public et ainsi légitimer leur recherche. C'est en raison de leur capacité à mettre des problèmes à l'agenda que les relations entre science et médias se seraient modifiées.

Plusieurs observations viennent éclairer cette nouvelle orientation des scientifiques vers les médias. Premièrement, les scientifiques accueillent et même recherchent le contact avec les médias et ces contacts font partie intégrante de leur activité (Peters 2012). Ils déclarent que les raisons pour lesquelles ils s'adressent aux médias sont majoritairement pour faire part de leur recherche au public (Ministère de l'Education Nationale) et ainsi essayer de

modifier l'opinion négative du public sur la science (Nisbet and Lewenstein 2002b). Ils indiquent également qu'ils souhaitent attirer l'attention des politiques et des financeurs sur leurs projets et favoriser leur carrière professionnelle (Allgaier et al. 2013; Peters 2013). D'ailleurs, plusieurs études ont indiqué que les chercheurs étaient plus enclins à interagir avec les médias que par le passé (Peters et al. 2008; Russell 1999). Deuxièmement, les instituts de recherche et les universités se sont dotés de services de communication et proposent maintenant des communiqués de presse qu'ils envoient aux journalistes (Peters et al. 2008; Schafer 2011b). Ils organisent aussi des conférences de presse pour présenter les découvertes réalisées en leur sein (Rodder 2009). Troisièmement, de leur côté, les agences de financements demandent dans les appels à projet de prévoir un paragraphe sur la dissémination publique des résultats de la recherche (Allgaier et al. 2013). Finalement, la *Royal Society* a publié un rapport en 2006 à la suite d'une enquête réalisée auprès de scientifiques britanniques. L'enquête portait sur leurs activités de « sensibilisation » du public, ou, autrement dit, leurs efforts pour communiquer leur recherche au public (The Royal Society 2006). On voit donc que les scientifiques et leurs institutions, de même que les journaux scientifiques, orientent leur activité, sinon directement vers le public, au moins vers les médias.

3.4- Les médias et l'incertitude

Plusieurs auteurs se sont intéressés à la façon dont les journalistes traitaient de l'incertitude lorsqu'ils présentaient des résultats scientifiques (Dudo et al. 2011; Friedman et al. 1999; Hornmoen 2010; Nelkin 1995). En général, les journalistes ont tendance à ne pas tenir compte de l'incertitude : ils ne mentionnent pas les limitations, simplifient les résultats et omettent le contexte (Lehmkuhl et al. 2016; Stocking et al. 2009). Ainsi, Dudo et al ont montré, dans leur analyse de la couverture médiatique des nanotechnologies entre 1988 et 2008, que l'incertitude liée à ces technologies n'est jamais mentionnée, que se soit pour en exprimer les risques potentiels ou pour en souligner le caractère émergent (Dudo et al. 2011). De même, une enquête d'opinion sur les nanotechnologies réalisée auprès de journalistes allemands a révélé qu'ils mentionnaient l'incertitude liée aux risques potentiels mais pas celle inhérente aux résultats de la recherche émergente (Guenther et al. 2013). Dans certains cas, à l'opposé, les journalistes choisissent d'exagérer l'incertitude : dans le cas de désaccords apparents entre scientifiques ou encore en exagérant les risques (Singer 1990). On peut penser par exemple à la couverture médiatique des risques d'autisme liés à la vaccination contre la rougeole, la rubéole et les oreillons (Clarke 2011) ou encore au traitement médiatique différent des technologies génétiques quand elles sont associées à la médecine (traitement positif) ou à l'alimentation (traitement négatif) (Eyck et al. 2003). Les journalistes peuvent

donc choisir de présenter la science comme plus ou moins certaine qu'elle ne l'est vraiment (Collins 1987; Friedman et al. 1999; Stocking et al. 2009). D'autres auteurs se sont intéressés aux motivations des journalistes à tenir compte de l'incertitude (Guenther et al. 2016; Lehmkuhl et al. 2016). Ainsi, l'intérêt des médias pour la controverse et les résultats conflictuels explique leur choix de traiter de l'incertitude : ils mentionnent l'incertitude quand elle est liée aux risques (Ruhrmann et al. 2015).

Si l'on se souvient des définitions de l'incertitude fournies par les sociologues des sciences proposées au début de ce chapitre (notamment celle de Callon et al. 2014), on voit que la seule incertitude dont il est question ici c'est celle liée au manque de connaissances. Les chercheurs ne savent pas encore et ou ne sont pas d'accord. Il n'existe pas encore de consensus parce que la recherche concerne une science émergente (applications et risques liés à la recherche sur les cellules souches, nanotechnologies) ou encore parce que les vues entre les scientifiques divergent. Encore une fois, l'incertitude inhérente à l'activité de recherche en train de se faire n'est pas prise en compte. Il semblerait donc que lorsqu'un résultat scientifique est publié dans un journal prestigieux, il soit synonyme de certitude. Le fait que le résultat soit nouveau, comme c'est souligné presque systématiquement dans les articles de presse, n'engage pas les journalistes à penser que ce résultat est donc préliminaire et qu'il doit être vérifié. Il est difficile de déterminer si ce « non-traitement » de l'incertitude liée à des résultats initiaux résulte de la méconnaissance des journalistes de la méthode scientifique ou s'ils choisissent tout simplement de l'ignorer.

Pourtant, l'incertitude liée aux résultats initiaux est mentionnée par plusieurs auteurs en sciences de la communication. Ainsi, dans l'ouvrage « *Communicating uncertainty : media coverage of new and controversial science* », Dunwoody distingue l'incertitude liée la science émergente –nouveaux résultats scientifiques ou nouvelles technologies- et la science controversée –désaccord entre les scientifiques ou résultats conflictuels dans la sphère publique- (Friedman et al. 1999). Cependant, à notre connaissance, à part l'étude réalisée sur l'hyperactivité (Gonon et al. 2012), il n'y a pas de recherche sur les conséquences de la couverture médiatique des résultats initiaux de la recherche biomédicale. Pourtant, ces résultats sont par définition incertains, ils ont tendance à être exagérés et sont souvent en désaccord avec les études ultérieures sur le même sujet. Mais, comme ils sont publiés dans des journaux prestigieux et accompagnés de communiqués de presse, ils attirent l'attention des journalistes. Et ceux-ci les présentent comme des faits avérés, donnant ainsi une information au public au mieux exagérée et au pire fausse.

Finalement, les sciences de la communication nous apportent un éclairage sur la présentation de sciences dans les médias. On voit que, dans la presse, les résultats sont exagérés et non conformes à la publication scientifique. Ce qui n'est pas gênant en soit, ce qui compte c'est que la compréhension du public n'en soit pas altérée (Hilgartner 1990). Si pendant longtemps les auteurs se sont intéressés au « fossé » qui opposait scientifiques et journalistes en raison des normes qui les séparent, il semble que celui-ci se soit considérablement réduit si tant est qu'il ait vraiment existé (Peters et al. 2008b). En effet, les scientifiques acceptent plus volontiers d'interagir avec les médias et sont prêts à accepter plus facilement des inexactitudes. D'autre part, l'incorporation de critères d'intérêt médiatique dans la sélection opérée par les journaux scientifiques prestigieux, les politiques de financement ou d'évaluation de la recherche, indiquent une orientation de l'activité scientifique vers les médias, concept décrit comme la *médialisation*.

IV- Notre thèse centrale

Le point de départ de cette thèse consiste à considérer que la crise de la reproductibilité des résultats de la recherche, notamment biomédicale, est un sujet d'inquiétude pour les scientifiques. Selon eux, elle sape les fondations de la recherche et questionne son intégrité. Elle ne concerne pas seulement la fraude ou la falsification des résultats, mais, plus profondément, la crise méthodologique qui frappe le monde de la recherche. En effet, soumis à des pressions de publication - contrepartie des politiques de recherche sur projet et des politiques d'évaluation - les chercheurs se laissent aller à des pratiques discutables, tant au niveau des publications scientifiques qu'au niveau de leurs interactions avec les médias. Si on utilise certains des cadres décrits dans cette introduction, on peut postuler que les normes proposées par Merton -plébiscitées par les scientifiques comme en témoigne leur inquiétude face à la « crise de la reproductibilité » - entrent en conflit avec les pratiques d'une recherche de type mode 2, finalisée, comme c'est le cas pour la recherche biomédicale. En effet, elle est porteuse de promesses. La Stratégie Nationale de Recherche (SNR), en proposant le « grand défi sociétal santé et bien-être », favorise l'extrapolation de résultats préliminaires - comme l'identification de facteurs de risques - à des perspectives de traitement de maladie. D'autre part, le concept de *médialisation* permet d'expliquer pourquoi les agences de financement demandent des indications de la diffusion des résultats de la recherche au grand public. On trouve également un bon exemple de cette *médialisation* dans la lettre adressée par l'Institut National des Sciences Biologiques du CNRS, à ses chercheurs. Il les enjoignait de « s'ouvrir à la société et (...) émerveiller nos concitoyens et les décideurs politiques (...) par la matière noire que représente la vie sur notre

planète, leur faire découvrir que la conquête du vivant est devant nous (...). Outre différentes manifestations, nous entreprenons la rédaction d'un livre blanc, qui se voudra le manifeste des sciences de la vie du XXIème siècle. » (Lettre de l'Institut National des Sciences Biologiques 2016).

Ces éléments semblent indiquer une évolution du mode de production des connaissances. L'hypothèse centrale de cette thèse est que les scientifiques, poussés par leurs institutions, les agences de financements de la recherche et les politiques d'évaluation ont tendance à exagérer l'importance et la portée de leurs résultats non seulement dans les publications scientifiques, dans leurs demandes de financement, mais aussi dans leurs interactions avec les médias. Ainsi, les contraintes institutionnelles et l'importance des médias dans la mise à l'agenda des problèmes et questions méritant un financement de recherche incitent les chercheurs à poursuivre des objectifs de financement et de publication plutôt qu'à « faire des sciences ». L'analyse de l'incertitude liée aux résultats initiaux dans les publications scientifiques, de son traitement dans les médias et sa perception par les journalistes étayera notre hypothèse.

En effet, il est par exemple fréquent de lire dans la presse que le sucre est bon pour la santé pour lire la semaine suivante qu'il est dangereux voire nocif d'en consommer trop. Comment arrive-t-on à de telles différences ? Comment est-il possible de lire des conclusions aussi diamétralement opposées ? Pour comprendre ces phénomènes, nous avons décidé d'analyser la reproductibilité des résultats de la recherche biomédicale, qui fait très souvent l'objet d'articles de presse. Nous avons donc créé une base de données de publications scientifiques s'intéressant à l'association entre des facteurs de risque (environnementaux, biologiques, génétiques) et des pathologies (somatiques, psychiatriques et neurologiques)¹¹. Nous avons identifié les études initiales et analysé leur reproductibilité en les comparant aux résultats de méta-analyses (Chapitre II). En effet, les études initiales sont les études qui revêtent le caractère le plus incertain mais qui sont néanmoins les plus visibles puisqu'elles sont souvent publiées dans des journaux prestigieux. Leur reproductibilité a été quantifiée en fonction de différents paramètres : les domaines de recherche, le type d'étude considéré (i.e. génétique, imagerie cérébrale), le prestige du journal scientifique dans lequel l'étude avait été publiée et l'année de publication. Ce premier chapitre nous permet d'évaluer ainsi l'incertitude présente au niveau des publications de la recherche biomédicale. Nous avons ensuite identifié, en utilisant cette même base de données, les études qui avaient reçu une

¹¹ Les précisions méthodologiques pour chaque composante de notre thèse seront livrées dans l'introduction de chacun des chapitres de la thèse.

couverture médiatique. Nous avons également analysé la reproductibilité des études couvertes et ainsi mesuré l'incertitude présente au niveau des articles de presse. Nous avons ainsi pu mettre en évidence certains des paramètres qui expliquaient le choix des études scientifiques relayées par les médias. Nous avons également analysé le contenu de ces articles de presse afin de déterminer la façon dont les journalistes décident de présenter, ou non, l'incertitude liée aux résultats initiaux de la recherche biomédicale (Chapitre III). Ces deux approches quantitatives nous permettent ainsi de mettre en lumière le degré d'incertitude des résultats de la recherche dans les journaux scientifiques d'une part et dans les articles de presse d'autre part. Pour comprendre le rôle des journalistes dans la sélection et la présentation des sujets, nous avons enfin utilisé une approche qualitative et réalisé des entretiens semi-directifs auprès de 21 journalistes couvrant des études scientifiques (Chapitre IV). A la lumière de ces résultats et de la littérature, nous proposons dans la dernière partie (Chapitre V) une analyse des interactions entre les sciences, les médias et les politiques de recherche ainsi que des pistes de réflexion et de futures recherches.

Cette thèse nous permet ainsi d'établir un lien entre les études réalisées dans différentes disciplines. En effet, l'incertitude inhérente à la pratique de la recherche est décriée dans les revues de sciences biologiques et médicales où les chercheurs s'indignent du manque de reproductibilité des résultats. Cependant, il nous semble que le problème n'est pas tant que les résultats ne soient pas reproductibles - c'est le fonctionnement normal de la recherche - mais plutôt qu'il y ait un excès de résultats positifs mis en avant dans les médias et une exagération volontaire de l'intérêt ou de la portée de ces résultats dans les publications scientifiques, mais aussi dans les demandes de financement. Si la sociologie des sciences s'intéresse à l'activité et à la production scientifique c'est bien souvent au travers de l'étude de controverses et l'incertitude intrinsèque à la recherche est largement ignorée pour se focaliser sur l'incertitude liée aux « états de monde possibles » (Callon et al. 2014). Pareillement, si les sciences de la communication se préoccupent de la façon dont les journalistes traitent de l'incertitude c'est plutôt dans les cas où l'incertitude est associée à des risques potentiels ou des sujets controversés.

Il semble donc qu'il n'y ait pas d'analyse des conséquences de l'incertitude liée aux résultats de la « science en train de se faire » sur la diffusion de ces mêmes résultats au public au travers de leur présentation par les médias. C'est ce que nous nous proposons d'explorer dans cette thèse.

CHAPITRE II

Incertitude et recherche - Analyse de la reproductibilité des résultats de la recherche biomédicale

De nombreux éditoriaux dans la littérature scientifique s'inquiètent du problème de reproductibilité qui frappe la recherche biomédicale. Leurs auteurs donnent leur opinion sur l'étendue de cette crise, ses causes et les façons d'y remédier (Alberts 2013; Ioannidis 2011, 2013; Munafo et al. 2014; Munafo et al. 2017; Sarewitz 2012). Ils insistent notamment sur leur inquiétude concernant le fonctionnement même de la recherche. En effet, cette médiocre reproductibilité n'est pas le fait de fraudes, qui sont rares et relativement faciles à identifier, mais de conduites discutables, moins identifiables. Ainsi, les biais de publication favorisent les résultats positifs qui seuls semblent attirer l'attention des journaux scientifiques. On se retrouve donc face à un excès de résultats positifs dans la littérature scientifique (Yong 2012). Comme nous l'avons décrit dans l'introduction (chapitre I), plusieurs facteurs entrent en jeu pour expliquer cette prépondérance des résultats positifs : la non-publication des résultats négatifs, les expériences réalisées sur de trop petits échantillons, la modification de l'hypothèse après obtention des résultats ou encore le changement de l'analyse statistique pour obtenir des résultats statistiquement significatifs. Pour les chercheurs en sciences biomédicales, cette « crise de reproductibilité » est un problème majeur qui affecte l'intégrité de la recherche et dont on ne connaît exactement ni l'étendue ni les conséquences. Néanmoins, le problème est suffisamment grave pour que la presse grand public s'en fasse l'écho (The Economist 2013a, 2013b). Les agences de financement de la recherche et les journaux scientifiques s'en préoccupent également et proposent des façons d'y remédier (Baker 2016b; Collins et al. 2014).

Dans un rapport écrit à l'issue d'un symposium sur la reproductibilité de la recherche biomédicale au Royaume-Uni, the *Academy of Medical Sciences* a donné une définition de la reproductibilité : un résultat est considéré comme reproductible quand d'autres équipes de recherche réalisent l'expérience décrite dans des conditions similaires et obtiennent des résultats comparables (Academy of Medical Sciences et al. 2015). Par conséquent, le résultat d'une seule étude ne saurait être synonyme de fait scientifique : il est par nature incertain et seule sa reproduction par d'autres chercheurs permettra sa validation. Il serait néanmoins illusoire de vouloir atteindre 100% de reproductibilité. Premièrement, certaines expériences sont difficilement répétables à l'identique même si les chercheurs ont correctement décrit et publié leurs méthodes. Deuxièmement, les procédures statistiques habituellement utilisées en biomédecine donnent l'illusion d'un degré de certitude plus élevé qu'en réalité. C'est ce

qu'explique très bien un article de la revue grand-public *The Economist* (The Economist 2013b). Sur la figure 1, les journalistes nous engagent à imaginer que sur 1000 hypothèses testées par les scientifiques, seulement 100 sont valides (points verts). Si les chercheurs testent ces 1000 hypothèses en utilisant un seuil de significativité statistique à $p < 0,05$, ceci implique que 45 des hypothèses testées seront des faux positifs (soit 5% des 900 hypothèses nulles indiquées en vert foncé). D'autre part si les chercheurs se placent dans des conditions de puissance statistique considérées comme satisfaisantes (0,8), ils ne pourront confirmer que 80 des 100 hypothèses valides et donc 20 résultats (en rouge) seront des faux négatifs. Finalement, en utilisant de tels standards, fréquemment utilisés dans la recherche biomédicale, 125 (80 vrais positifs et 45 faux positifs) des hypothèses testées seront déclarées positives. Quarante cinq études sur 125 (36%) annoncent donc des résultats positifs qui sont en réalité de fausses découvertes. On remarquera qu'à l'inverse les résultats négatifs sont bien plus fréquemment conformes à la réalité : seulement 45 des 900 hypothèses nulles ont été incorrectement identifiées comme non nulles. Malheureusement, la majorité des articles publiés en biomédecine rapportent des résultats significativement en accord avec l'hypothèse testée. Le pourcentage d'hypothèses validées ne cesse d'augmenter : il est passé de 70% en 1990 à 88% en 2007 (de Winter et al. 2015; Fanelli 2012).

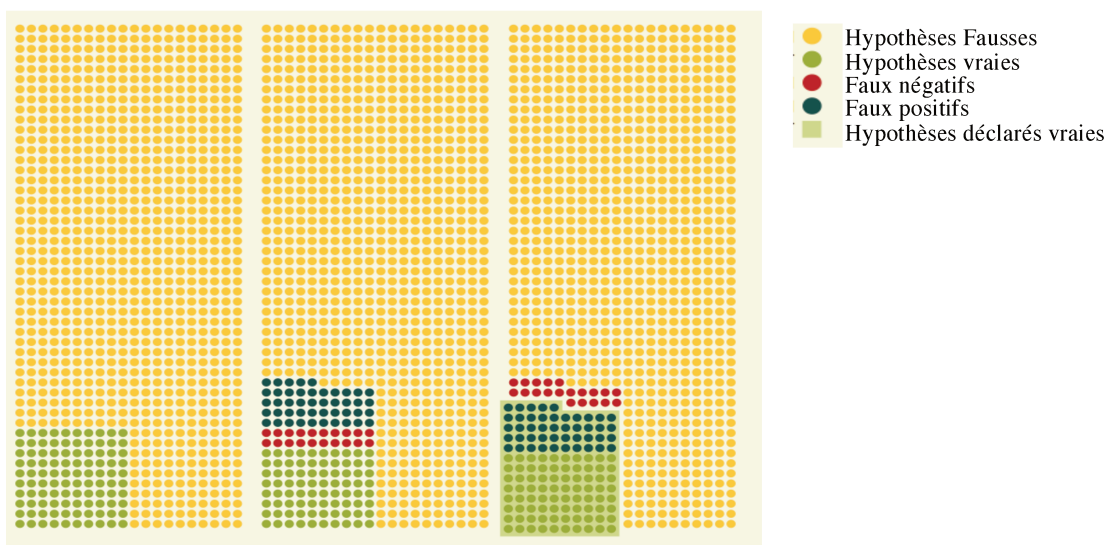


Figure 1 : Fausses découvertes. D'après *The Economist* (The Economist 2013b)

Ce problème est exacerbé lorsque les échantillons utilisés sont de trop petite taille ce qui diminue la puissance statistique (Ioannidis 2005). D'autre part, comme la plupart des associations facilement détectables (comme par exemple l'association entre une mutation dans le gène CFTR et la mucoviscidose) ont déjà été identifiées, les scientifiques recherchent

maintenant des associations plus ténues et plus difficiles à détecter. Les tailles d'effet¹² sont donc plus petites et ont plus de chances d'être invalidées par des études ultérieures (Ioannidis 2005]. En effet, si les « odd ratios » (OR)¹³ estimés sont autour de 1,05, comme c'est le cas pour de nombreux facteurs de risques génétiques, non seulement leur valeur informative intrinsèque est faible (un OR de 1 signifie qu'il n'y a pas d'association entre les facteurs), mais en plus, les associations considérées comme significatives ne le sont souvent pas.

Dans le contexte actuel de la recherche, où la prime est donnée aux résultats positifs et nouveaux, les réplifications des études sont relativement rares. En effet, après la publication initiale d'un résultat significatif, une deuxième équipe aura du mal à publier dans une revue prestigieuse. De plus, une deuxième étude contredisant la première par un résultat négatif devra, pour être convaincante, s'appuyer sur un nombre d'observations beaucoup plus élevé afin d'exclure de son intervalle de confiance la taille d'effet rapportée comme significative par la première étude. Par conséquent, une étude de réplification convaincante présente un coût économique plus élevé qu'une étude initiale.

Au total deux facteurs principaux concourent à la « crise de la reproductibilité » : la publication préférentielle de résultats positifs en accord avec l'hypothèse de départ et les contraintes décourageant les études de réplification. Notons ici que ces deux facteurs agissent même dans les conditions normales du fonctionnement actuel de la recherche. Ils ne tiennent pas compte de la fraude caractérisée, qui est rare, ni des pratiques douteuses, qui sont beaucoup plus fréquentes.

Plusieurs publications se sont attachées à quantifier la reproductibilité des résultats de la recherche biomédicale, que ce soit dans le cas des études précliniques, des essais cliniques et des études d'association¹⁴. Ainsi, dans un article récent, des auteurs ont analysé la reproductibilité des études précliniques (Freedman et al. 2015). Ils ont montré, en collectant les données de 5 études différentes, qu'entre 51 et 89% des études précliniques n'étaient pas reproduites. Dans le cas des essais cliniques le constat est le même : 60% des essais cliniques de phase II présentant un effet bénéfique ne sont pas confirmés lors des phase III (DiMasi et al. 2010). Pareillement, la majeure partie des études initiales sur un nouveau traitement rapporte des tailles d'effet bien supérieures à celles observées dans les études ultérieures ou

¹² Taille d'effet : mesure quantitative de la force de la relation entre deux variables.

¹³ *Odd ratio* : mesure statistique qui permet d'évaluer le degré de dépendance entre deux variables. Par exemple être porteur d'un gène et développer une maladie

¹⁴ Les études d'association sont des études qui mesurent la corrélation (association) entre un facteur de risque et une maladie.

les méta-analyses¹⁵ (Ioannidis et al. 2001; Ioannidis 2005b; Pereira et al. 2012; Tajika et al. 2015; Trikalinos et al. 2004). Cette inflation des résultats initiaux a aussi été observée dans le cas des études d'association, notamment en génétique : les études initiales présentent des tailles d'effet plus grandes que celles des études ultérieures (Hirschhorn et al. 2002; Ioannidis et al. 2001b; Trikalinos et al. 2004b). Les mêmes problèmes sont retrouvés dans les études d'association entre des marqueurs biologiques et des pathologies (Ioannidis et al. 2011c). Ainsi, Ioannidis et al, ont comparé les tailles d'effet rapportées par 35 articles scientifiques largement cités dans la littérature (plus de 400 fois) avec celles présentées dans les méta-analyses : dans 83% des cas, les tailles d'effet des études hautement citées leurs étaient supérieures. Les études largement citées ne sont donc pas plus valides. Si l'on restreint l'analyse aux études initiales (14 études sur 35 articles) le taux de non reproductibilité est de 60% : une des études a été invalidée par la méta-analyse et sept autres présentaient des tailles d'effet au moins deux fois plus élevées que celles mesurées dans les méta-analyses (Ioannidis et al. 2011c).

Tout cela étant dit, les études présentées ci-dessus ont été réalisées sur un nombre limité d'observations. Il nous a semblé important de quantifier plus largement la reproductibilité des études initiales de la recherche biomédicale. Pour réaliser ce projet, nous avons décidé d'utiliser des méta-analyses. En effet, elles constituent un outil de choix. Elles compilent les résultats de plusieurs études sur un même sujet et donnent ainsi une estimation de la taille d'effet basée sur un large nombre d'observations. D'autre part, elles ont été largement utilisées par l'équipe de Ioannidis (Fanelli et al. 2017; Ioannidis et al. 2001; Ioannidis 2005b; Pereira et al. 2012; Trikalinos et al. 2004).

Nous avons choisi d'étudier la reproductibilité des études d'association entre des facteurs de risques et des pathologies. Ce type de recherche est particulièrement important puisqu'il permet d'améliorer la prévention, le diagnostic et éventuellement le traitement de ces pathologies. Nous n'avons pas étudié les essais cliniques testant l'efficacité des traitements des maladies puisque ceux-ci sont souvent financés par l'industrie pharmaceutique. Comme nous l'avons décrit dans l'introduction, les conflits d'intérêt inévitables risquent d'aggraver dans ce cas l'exagération des résultats des études initiales.

Nous avons sélectionné des méta-analyses associant différents facteurs de risque comme des risques génétiques, biochimiques ou environnementaux avec douze pathologies réparties dans 3 domaines : la psychiatrie, la neurologie et d'autres maladies somatiques.

¹⁵ Une méta-analyse est une analyse statistique combinant les résultats d'une série d'études indépendantes sur un même sujet pour en faire la synthèse. Le résultat ainsi obtenu est considéré comme donnant une estimation de la taille d'effet la plus proche de l'effet réel.

Nous avons choisi 4 pathologies pour chacun de ces domaines : le trouble déficitaire de l'attention avec hyperactivité (TDAH), la schizophrénie, l'autisme et la dépression unipolaire pour la psychiatrie, l'épilepsie, la sclérose en plaque et les maladies d'Alzheimer et de Parkinson pour la neurologie. Nous avons aussi sélectionné 4 autres maladies somatiques, le cancer du sein, la polyarthrite rhumatoïde, le glaucome et le psoriasis. Pour chaque méta-analyse, nous avons identifié l'étude initiale, c'est à dire la première étude publiée sur le sujet, et l'étude la plus large, c'est à dire celle incluant le plus grand nombre de patients. L'objectif de cette étude était de (1) quantifier la reproductibilité des études initiales de plusieurs domaines de la recherche biomédicale et de (2) mesurer l'influence de certains facteurs, tels que le facteur d'impact du journal ou l'année de publication, sur la reproductibilité. Cette étude a été publiée en juin 2016 dans le journal PLoS ONE et a été réalisée et écrite en collaboration avec une équipe britannique de l'université de Bristol (Dumas-Mallet et al. 2016). Elle est incluse dans l'annexe 1. Les méthodes et outils d'analyse utilisés pour cette recherche sont décrits dans le paragraphe suivant. Les résultats sont ensuite présentés et discutés dans les paragraphes II et III. Dans le dernier paragraphe (IV) nous présentons une synthèse de ces résultats à la lumière de la littérature et proposons des éclairages sur « la crise de la reproductibilité ».

I- Méthodologie

Toutes les données relatives à cette recherche (étapes de la sélection des différentes méta-analyses pour chaque pathologie, tableaux *excel* répertoriant les différentes données extraites des publications) sont présentées dans l'annexe 1.

1.1- Sélection des méta-analyses

Nous avons utilisé la base de données PubMed pour identifier les méta-analyses d'intérêt. Nous avons limité la recherche aux méta-analyses rédigées en anglais et publiées entre le 1 janvier 2008 et le 31 décembre 2012. Certains journaux scientifiques proposent une publication dématérialisée avant la publication sous forme papier. Ainsi, si les articles étaient publiés sous forme écrite en 2013 ou 2014 mais publiés en ligne en 2012, ils ont été inclus dans la base de données. Nous avons utilisé des mots clés pour chaque pathologie : "*attention deficit hyperactivity disorder*", "*autism*", "*major depressive disorder*", "*epilepsy*", "*Alzheimer disease*", "*Parkinson disease*", "*multiple sclerosis*", "*breast cancer*", "*glaucoma*", "*psoriasis*" et "*rheumatoid arthritis*". Les titres et résumés des méta-analyses ainsi obtenues ont tous été lus indépendamment par deux auteurs (EDM et FG) et les articles complets ont été collectés. Les méta-analyses retenues devaient répondre à plusieurs critères :

elles devaient (1) compiler les résultats d'au moins sept études primaires issues d'au moins quatre publications indépendantes, (2) présenter les résultats sous forme d'*odd ratio* (OR), de risque relatif (RR), de différence en moyenne standardisée (*d* de Cohen) ou d'écart moyen¹⁶ et (3) ne pas concerner le traitement, dépistage ou diagnostic des pathologies considérées. D'autre part, les articles identifiés par cette recherche ont été exclus si (1) ils ne présentaient en fait pas de résultats de méta-analyses, (2) ils exprimaient les résultats sous une autre forme que définie ci-dessus, (3) ils n'étaient pas exploitables (données manquantes ou sans rapport direct avec la pathologie) ou encore (4) lorsque le facteur de risque analysé était la pathologie elle-même (par exemple la dépression considérée comme facteur de risque de développer des ulcères).

Dans le cas où plusieurs méta-analyses étaient présentées dans une même publication, elles ont toutes été sélectionnées si elles étudiaient des paramètres différents. Si les paramètres n'étaient pas différents, les méta-analyses incluant le plus grand nombre de données ont été sélectionnées. Enfin, lorsque plusieurs méta-analyses publiées dans des articles différents avaient étudié le même paramètre, nous avons choisi la méta-analyse la plus récente.

1.2- Sélection des études initiales et des études les plus larges

Pour chacune des méta-analyses ainsi sélectionnées nous avons identifié l'étude initiale et récupéré la publication dans laquelle cette étude était présentée. Chacune de ces publications a été lue afin de s'assurer que l'étude identifiée comme initiale dans la méta-analyse était réellement la première étude réalisée sur le sujet. Nous avons ainsi identifié et inclus dans la base de données 110 études initiales qui n'avaient pas été prises en compte par les méta-analyses correspondantes. Nous avons aussi identifié pour chaque méta-analyse l'étude la plus large, c'est à dire celle qui comprenait le plus grand nombre de patients.

1.3- Extraction des données

L'extraction des données a été réalisée indépendamment par deux auteurs (EDM et FG) et les éventuels désaccords ont été résolus après discussion. Pour s'assurer de l'exactitude des données extraites, un troisième auteur (KB) a contrôlé, de manière aléatoire, 10% des données. Très peu d'erreurs ont ainsi été détectées et ont été immédiatement corrigées.

Les données suivantes ont été extraites pour chaque publication (étude initiale, étude la plus large et méta-analyse) : l'année et le nom du premier auteur de la publication, la taille

¹⁶ Le risque relatif mesure le risque de survenue d'un événement dans un groupe par rapport à un autre. La différence en moyenne standardisée (*d* de Cohen) et les écarts moyens sont des mesures de taille d'effet.

d'effet, sa signifiante statistique ainsi que la valeur de p lorsque le résultat était significatif ou l'intervalle de confiance à 95% lorsque le résultat était non significatif, le nombre de patients inclus dans l'étude, le facteur d'impact du journal (déterminé en 2012) dans lequel l'étude a été publiée et, uniquement pour les méta-analyses, le nombre d'études primaires incluses dans la méta-analyse.

Les tailles d'effet exprimées en écart moyen ont été converties en d de Cohen en utilisant le logiciel OpenMetaAnalyst (Wallace et al. 2009). Toutes les tailles d'effet ont été converties de manière à pouvoir être comparées. Ainsi les d de Cohen ont été convertis en $\ln(\text{OR})$ en utilisant la formule $\ln(\text{OR}) = d \times \pi \sqrt{3}$. Pour chaque pathologie, nous avons également calculé la taille d'effet moyenne ainsi que le nombre moyen de patients pour les études initiales, les plus larges et les méta-analyses.

1.4- Sélection de méta-analyses robustes

Dans un premier temps, nous avons utilisé l'ensemble des 663 méta-analyses pour nos analyses. Dans un deuxième temps, nous avons procédé à une analyse de sensibilité en nous limitant aux méta-analyses considérées comme suffisamment robustes. Une méta-analyse robuste rapportant un effet significatif a été définie comme robuste si $p < 0,005$. Une méta-analyse statistiquement non significative (n'écartant pas l'hypothèse nulle) a été considérée comme robuste lorsque l'intervalle de confiance était suffisamment étroit pour exclure même un petit effet. La valeur de ces petites tailles d'effet a été calculée pour chaque facteur de risque étudié. Pour cela, nous avons distribué les tailles d'effet des méta-analyses en quintiles. Nous avons défini la plus petite taille d'effet (le seuil) comme étant la valeur de la taille d'effet la plus grande dans le dernier quintile. Pour les tests comportementaux, les 65 méta-analyses significatives ont permis de définir un seuil à $d \leq 0,51$. Pour les tests d'imagerie cérébrale, les 69 méta-analyses statistiquement significatives étaient toutes exprimées en d de Cohen, sauf deux exprimées en *odd ratio*. Après conversion, une taille d'effet de petite taille a été calculée : $d \leq 0,29$. Dans le cas des facteurs génétiques, il y avait 136 méta-analyses significatives et le seuil a été défini à $\text{OR} \leq 1,11$. Enfin, pour les autres facteurs de risque, les 146 méta-analyses significatives ont été séparées en deux groupes, 60 exprimées en d de Cohen ont permis de définir un seuil à $d \leq 0,3$ et les 96 exprimées en *odd ratio* ont permis de définir un seuil à $\text{OR} \leq 1,20$.

1.5- Analyse de la sensibilité et de la spécificité

Comme d'autres auteurs, nous avons utilisé les outils statistiques de spécificité et de sensibilité pour comparer les résultats des études initiales à ceux des méta-analyses (Trikalinos et al. 2004b). Ces outils sont généralement utilisés dans le cadre des tests

diagnostic. La sensibilité est ainsi définie : elle mesure la capacité d'un test à conclure à un résultat positif lorsqu'une hypothèse est vérifiée, c'est à dire à distinguer les faux négatifs. La spécificité elle mesure la capacité d'un test à conclure à un résultat négatif lorsque l'hypothèse n'est pas vérifiée, c'est à dire à s'affranchir des faux positifs. Ainsi la sensibilité d'un test est de 100% lorsque ce test est capable de détecter avec raison toutes les personnes malades et la spécificité est de 100% lorsqu'aucune personne saine n'est diagnostiquée malade à tort. Dans le cas de notre étude, la sensibilité est la probabilité qu'une étude initiale présente un résultat statistiquement significatif et que la méta-analyse correspondante confirme cette valeur nominale significative. La spécificité est la probabilité que l'étude initiale et la méta-analyse soient en accord pour ne pas rejeter l'hypothèse nulle. Il y a donc quatre possibilités d'accord ou non entre les méta-analyses et les études initiales : (1) l'étude initiale et la méta-analyse rapportent toutes les deux des effets statistiquement significatifs, (2) l'étude initiale et la méta-analyse rapportent toutes les deux des effets non significatifs, (3) l'étude initiale rapporte un effet statistiquement significatif contrairement à la méta-analyse (faux positif) et (4) l'étude initiale ne rapporte pas d'effet significatif au contraire de la méta-analyse (faux négatif). Pour six études initiales, nous avons observé un cas différent : méta-analyses et études initiales présentaient toutes les deux des résultats statistiquement significatifs mais dans des directions opposées (par exemple l'étude initiale indiquait qu'une région cérébrale était plus large dans le cas de la schizophrénie alors que la méta-analyse indiquait qu'elle était plus petite). Ces six études ont été considérées comme des faux positifs. La sensibilité et la spécificité ont été calculées avec « *the clinical research calculator 1* » (vassarstats.net).

II- Résultats

2.1- Caractéristiques des études incluses dans la base de données

Nous avons collecté 663 méta-analyses s'intéressant aux associations entre différents facteurs de risque et 12 pathologies, 4 pour chaque domaine de recherche, la psychiatrie, la neurologie et les maladies somatiques.

Nous avons différencié 4 types de facteurs : (1) les facteurs cognitifs et les tests comportementaux, (2) les résultats d'imagerie cérébrale, (3) les facteurs génétiques et (4) les autres facteurs qui couvrent une grande variété de facteurs différents (risques épidémiologiques, tests sanguins et biochimiques, caractéristiques physiques).

Les facteurs (1) et (2) concernent presque exclusivement le domaine de la psychiatrie. Les autres types de facteurs (3 et 4) sont retrouvés de manière similaire pour les 3 domaines.

Les résultats des différentes étapes de la sélection des études incluses dans notre base de donnée sont présentés dans tableau 1.

Nous avons également identifié les méta-analyses robustes. En effet des études ont montré que si une méta-analyse était statistiquement significative à $p < 0,05$, elle pouvait être contredite par des méta-analyses ultérieures incluant un plus grand nombre d'observations. Par contre, lorsque le seuil de significativité était à $p < 0,005$, elles étaient plus rarement contredites (Ioannidis 2008b; Pereira et al. 2011).

Tableau 1 : nombre d'études identifiées dans chacun de 3 domaines. TDAH : trouble de déficit de l'attention avec hyperactivité, Aut : autisme, Dep : dépression, Sch : schizophrénie, MA : maladie d'Alzheimer, Epi, épilepsie, SP, Sclérose en plaque, MP : maladie de Parkinson, CS : cancer du sein, Glau : glaucome, Pso : psoriasis, PR : polyarthrite rhumatoïde.

	Pathologies psychiatriques				Maladies neurologiques				Maladies somatiques			
	TDAH	Aut	Dep	Sch	MA	Epi	SP	MP	CS	Glau	Pso	PR
Recherche Pubmed	118	71	553	454	197	147	137	139	811	75	81	285
Articles examinés	43	36	114	198	117	29	44	74	345	19	22	82
Articles inclus	20	13	28	59	35	12	18	20	89	9	8	22
Méta-analyses incluses	40	24	53	203	50	15	37	57	114	21	15	34
Tests cognitifs/ comportementaux 1	14	0	9	43	1	0	0	1	0	0	0	0
Imagerie cérébrale 2	2	10	19	64	1	1	0	0	0	0	0	0
Facteurs génétiques 3	12	4	8	56	33	3	17	30	61	17	12	31
Autres facteurs 4	12	10	17	40	15	11	20	26	53	4	3	3
Méta-analyses robustes	24	11	36	134	20	9	19	37	57	10	8	20

La sélection des méta-analyses statistiquement significatives à $p < 0,005$ permettrait ainsi d'augmenter la validité du résultat (Johnson 2013). D'autre part, un résultat non significatif implique que l'hypothèse nulle n'a pas pu être rejetée, cela ne prouve pas l'absence d'association. Néanmoins, il est possible que l'hypothèse nulle n'est pas pu être rejetée parce que, finalement, il n'y a pas d'association entre les deux paramètres mesurés. Ainsi, en utilisant les valeurs de l'intervalle de confiance à 95%, il est possible de définir un effet seuil au delà duquel il sera possible de conclure que la probabilité que deux paramètres soient associés est négligeable (Dienes 2014). La difficulté est alors de définir ce seuil, la taille d'effet minimale en dessous de laquelle il est possible de conclure qu'il n'y a pas d'association entre deux paramètres. En effet, les tailles d'effet moyennes changent en fonction des facteurs d'association étudiés (Ioannidis 2008b). Comme expliqué dans la section méthodologie, nous avons utilisé les tailles d'effet des 425 méta-analyses statistiquement significatives pour déterminer la taille d'effet la plus petite pour chacune des catégories de facteurs (tests cognitifs, imagerie cérébrale, tests génétiques et autres tests). Nous avons ainsi identifié 385 méta-analyses robustes dont 326 étaient statistiquement significatives.

2.2- Accord statistique entre les études initiales et les méta-analyses

Comme nous l'avons expliqué dans le paragraphe 1.5, nous avons calculé la sensibilité et la spécificité des études initiales ou, autrement dit, leur capacité à détecter les

faux négatifs et les faux positifs. Les résultats sont présentés dans le tableau 2. La sensibilité des études initiales est comprise entre 0,58 et 0,64 et est relativement similaire quel que soit le domaine étudié. Par contre, la spécificité diffère en fonction des domaines et varie entre 0,47 et 0,7. Elle est plus faible en psychiatrie si on la compare à celle de neurologie (χ^2 test: $p = 0.0055$) ou à celle des autres maladies somatiques (χ^2 test: $p = 0.059$). Par conséquent, les études initiales en psychiatrie semblent générer plus de faux positifs.

Nous avons représenté ces résultats sous la forme d'un diagramme ROC. Il exprime la sensibilité (vrais positifs) en fonction de 1-spécificité (faux positifs). La figure 2 montre que dans le cas de la psychiatrie, la probabilité pour que le résultat de l'étude initiale soit en accord avec celui de la méta-analyse n'est pas meilleure que celle obtenue au hasard.

Si l'on considère seulement les méta-analyses robustes les résultats sont similaires (voir tableau 2 et diagramme ROC figure 2).

Tableau 2 : Sensibilité et spécificité des résultats études initiales comparées à ceux des méta-analyses. Les intervalles de confiance sont ceux à 95%.

	Pathologies psychiatriques	Maladies neurologiques	Maladies somatiques
Méta-analyses (toutes)	320	159	184
# statistiquement significatives	221 (69,1 %)	100 (62,9 %)	104 (56,5 %)
Sensibilité	0,612 (0,544; 0,676)	0,636 (0,533; 0,729)	0,584 (0,482; 0,680)
Spécificité	0,475 (0,376; 0,577)	0,700 (0,566; 0,808)	0,614 (0,501; 0,717)
Méta-analyses robustes	205	85	95
# statistiquement significatives	178 (86,8 %)	77 (90,6 %)	71 (74,7 %)
Sensibilité	0,625 (0,549; 0,696)	0,649 (0,531; 0,752)	0,718 (0,597; 0,816)
Spécificité	0,533 (0,346; 0,712)	0,875 (0,477; 0,993)	0,625 (0,408; 0,804)

2.3- Les tailles d'effet des études initiales sont elles exagérées ?

L'accord nominal entre les études initiales et les méta-analyses n'implique pas que les valeurs de ces tailles d'effet soient similaires. En effet, nous avons vu que les tailles d'effet rapportées dans les études initiales étaient souvent supérieures à celles rapportées dans les méta-analyses (Ioannidis 2005b). Nous avons donc comparé les tailles d'effet des études initiales avec celles des méta-analyses. Pour les 256 études initiales rapportant un effet statistiquement significatif et appariées à une méta-analyse également significative, on constate que 150 d'entre elles, soit 59%, présentent une taille d'effet exagérée de plus de 50% par rapport à celle des méta-analyses, et que 97 d'entre elles rapportaient un effet exagéré de plus de 100%.

Si l'on représente les tailles d'effet moyennes des études initiales pour chaque pathologie en fonction de celles de leur méta-analyse correspondante, on constate que les études initiales rapportent en moyenne toutes des tailles d'effet exagérées. En effet, elles se situent toutes au dessus de la droite d'équivalence qui représente l'accord entre les résultats des méta-analyses et ceux des études initiales. Nous avons observé cet écart aussi bien en considérant l'ensemble des études initiales (Figure 3A) qu'en limitant l'analyse aux 359 études initiales significatives (Figure 3C). Pour chaque pathologie, la distribution des tailles d'effet des études initiales diffère de celle des méta-analyses correspondantes. En utilisant le test de Wilcoxon pour les données appariées, nous avons ainsi montré que cette différence atteignait un haut degré de significativité statistique ($p < 0,002$) sauf pour l'autisme ($p = 0,025$), l'épilepsie ($p = 0,02$) et le psoriasis ($p = 0,009$).

Par contre, si l'on compare les tailles d'effet des études les plus larges (i.e. avec le plus grand nombre de patients) à celles des méta-analyses correspondantes, l'accord entre les tailles d'effet est bien supérieur (Figure 3B et 3D).

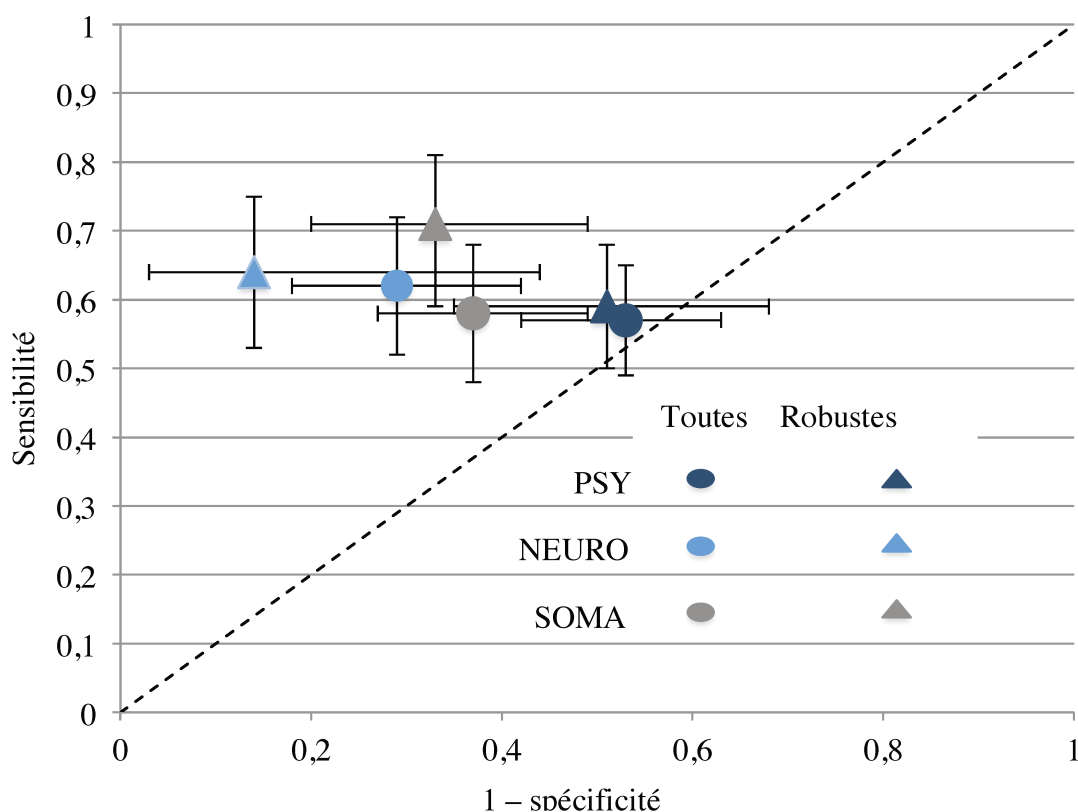


Figure 2 : Diagramme ROC : il représente la probabilité que le résultat des études initiales soit en accord avec celui des méta-analyses, que l'effet soit statistiquement significatif ou non. Nous avons représenté les résultats pour toutes les méta-analyses (663 représentées par des cercles) ou seulement pour les méta-analyses robustes (385 représentées par des triangles). Nous les avons distribuées en fonction de chacun des domaines : PSY pour les maladies psychiatriques, NEURO pour les maladies neurologiques et SOMA pour les maladies somatiques. Les barres d'erreurs correspondent aux valeurs de l'intervalle de confiance à 95%. La ligne en pointillé correspond à des accords qui auraient pu être obtenus de manière purement aléatoire. Les barres d'erreur pour les maladies psychiatriques

chevauchent cette ligne. Ceci suggère donc que dans le cas des études initiales en psychiatrie, la probabilité qu'elles soient en accord avec les méta-analyses est aléatoire.

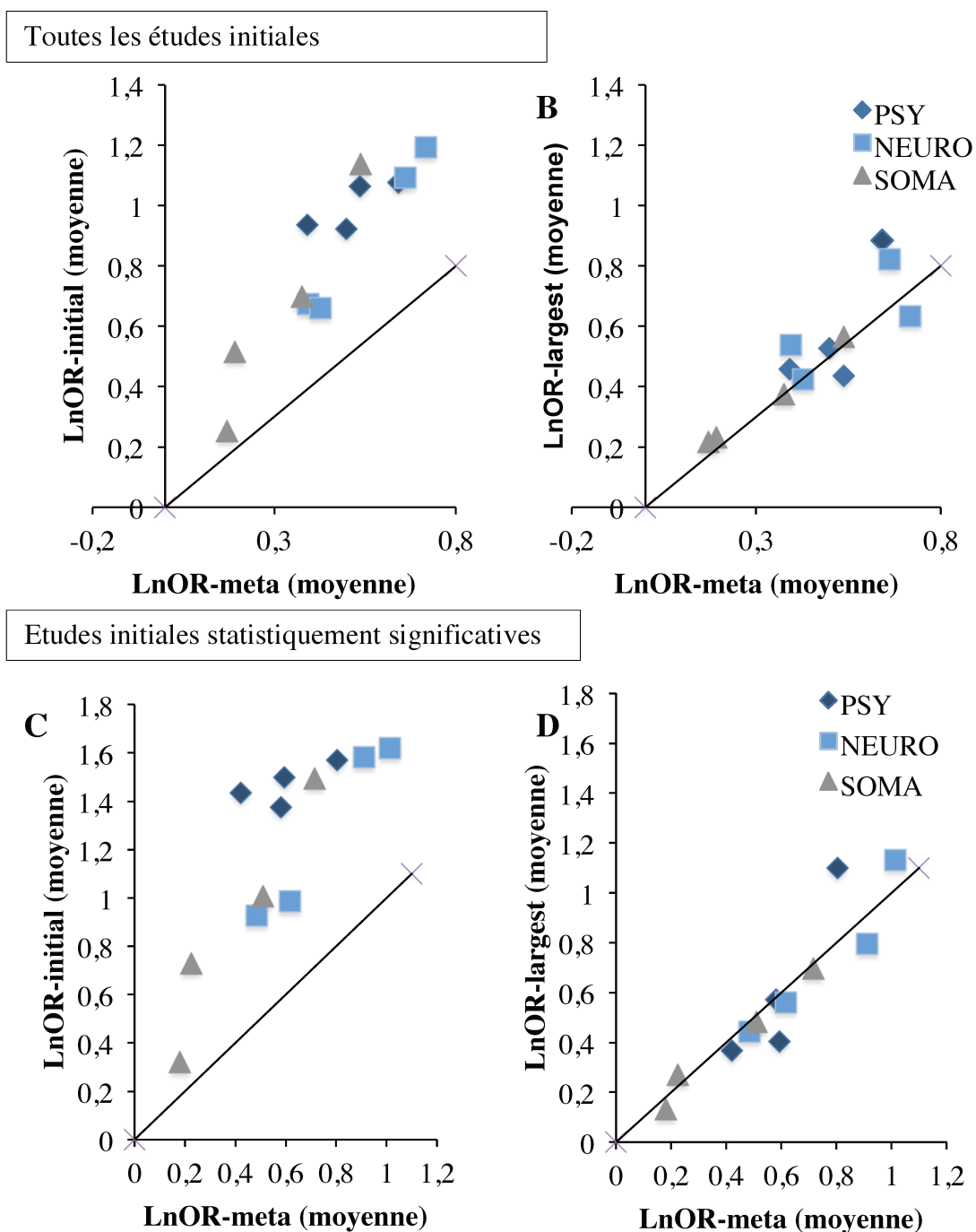


Figure 3 : Comparaison entre les tailles d'effet rapportées par les études initiales et les études les plus larges avec celles des méta-analyses correspondantes. Toutes les tailles d'effet ont été converties en Ln(OR). Les graphiques du haut (A et B) ont été réalisés à partir des données de toutes les études initiales, de leur études les plus larges et des méta-analyses correspondantes. Les graphiques du bas (C et D) ont été réalisés à partir des données des études initiales statistiquement significatives uniquement. Les graphiques A et C représentent les moyennes des tailles d'effet des études initiales pour chaque pathologie en fonction des moyennes des tailles d'effet des méta-analyses correspondantes. Les graphiques B et D représentent les moyennes des tailles d'effet des études les plus larges en fonction de celles des méta-analyses correspondantes.

2.4- Taux de reproductibilité des études initiales

2.4.1- En fonction des différents domaines : psychiatrie, neurologie et maladies somatiques

Nous avons considéré qu'une étude initiale était reproduite (1) si elle rapportait un effet non significatif comme la méta-analyse correspondante ou (2) si elle rapportait un effet statistiquement significatif comme la méta-analyse correspondante ne dépassant pas un certain seuil d'inflation. Nous avons testé deux valeurs de seuil : 50% ou 100% d'inflation (Tableau 3). Nous avons aussi mesuré ce taux de reproductibilité sur plusieurs sous-populations : toutes les études initiales, seulement les études initiales significatives et seulement les études initiales hautement significatives ($p < 0,005$). Les résultats sont aussi représentés sous forme d'histogrammes (Figure 4). On constate que, quel que soit le domaine de recherche biomédical analysé, le taux de reproductibilité est faible et varie entre 36,3 et 53,5%. D'autre part, ce taux de reproductibilité ne varie pas lorsqu'on restreint l'analyse aux études initiales hautement significatives à $p < 0,005$. Donc, même des études initiales avec une valeur de p élevée ne sont pas plus reproductibles que des études initiales avec un p compris entre 0,05 et 0,005.

Nous avons poursuivi notre analyse avec le seuil d'inflation à 100% puisque les résultats sont assez similaires si l'on n'utilise un seuil à 50% (Tableau 3). Avec un seuil à 100%, le nombre d'études initiales considérées comme confirmées est aussi plus grand et permet une meilleure analyse statistique.

En ce qui concerne les pathologies psychiatriques, on constate également que le taux de reproductibilité des études initiales considérées dans leur totalité est inférieur à celui des études initiales en neurologie ($p = 0,0094$). Il n'existe pas de différence significative entre les autres domaines et cette différence disparaît lorsque l'analyse est restreinte aux études initiales statistiquement significatives (à $p < 0,05$ ou $p < 0,005$).

Tableau 3 : Etudes initiales en accord avec les méta-analyses

	Nombre d'études						Comparaisons		
	PSY		NEURO		SOMA		PSY/	PSY/	NEURO/
	#	%	#	%	#	%	NEURO	SOMA	SOMA
Toutes les études initiales	320		159		184				
Accord et inflation $\leq 100\%$	131	40,9	85	53,5	84	45,7	$p=0,0094$	$p=0,3$	$p=0,15$
Accord et inflation $\leq 50\%$	101	31,6	71	44,7	75	40,8	$p=0,0049$	$p=0,037$	$p=0,47$
Initiales significatives	187		81		91				
Accord et inflation $\leq 100\%$	83	44,4	43	53,1	33	36,3	$p=0,19$	$p=0,2$	$p=0,27$
Initiales significatives $p < 0,005$	94		44		48				
Accord et inflation $\leq 100\%$	42	44,7	23	52,3	23	47,9	$p=0,41$	$p=0,71$	$p=0,68$

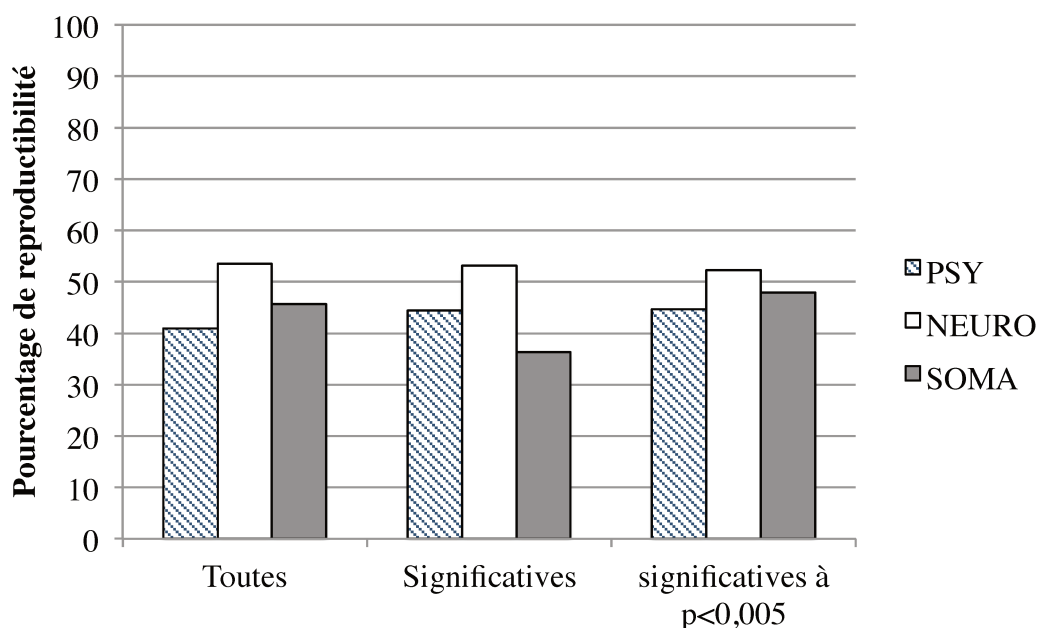


Figure 4 : Taux de reproductibilité des études initiales en fonction des différents domaines et de sous-populations.

Nous avons également mesuré le taux de reproductibilité des études initiales en limitant l'analyse aux méta-analyses robustes (Tableau 4). Là encore les résultats sont comparables : le taux de reproductibilité varie entre 45,4% et 72,5%.

Tableau 4 : taux de reproductibilité des études initiales limité aux méta-analyses robustes

	Nombre d'études						Comparaisons		
	PSY		NEURO		SOMA		PSY/ NEURO	PSY/ SOMA	NEURO/ SOMA
	#	%	#	%	#	%			
Toutes les initiales	205		85		95				
Accord et inflation ≤100%	93	45,4	44	51,8	47	49,5	p=0,32	p=0,51	p=0,76
Initiales significatives	124		51		60				
Accord et inflation ≤100%	77	62,1	37	72,5	32	53,3	p=0,19	p=0,26	p=0,037

2.4.2- En fonction du type d'association étudié

Nous avons ensuite analysé le taux de reproductibilité des études initiales en fonction du facteur de risque étudié et du domaine de recherche. Nous avons distingué 4 types de facteurs : les tests cognitifs/comportementaux, les résultats d'imagerie cérébrale, les facteurs génétiques et les autres facteurs. Nous avons réparti les études initiales en fonction de ces différents facteurs et en fonction du domaine biomédical étudié. Les facteurs tests cognitifs et études d'imagerie cérébrale ne sont pris en compte que pour le domaine de la psychiatrie. En effet, aucun de ces deux types de facteurs n'est étudié pour les maladies somatiques et, pour les maladies neurologiques, seules deux méta-analyses étudient des tests cognitifs et seulement quatre des résultats d'imagerie cérébrale. Par conséquent, nous avons analysé les 4 types de facteurs en association avec les troubles psychiatriques mais seulement les facteurs génétiques et les autres facteurs pour les maladies somatiques et la neurologie (Tableau 5).

Tableau 5 : Reproductibilité des études initiales en fonction des différents domaines et des types de facteurs de risque étudiés. Une étude est considérée comme reproduite si elle est en accord avec la méta-analyse et que le taux d'inflation du résultat de l'étude initiale n'excède pas 100%. PSY : Psychiatrie, NEURO : neurologie, SOMA : maladies somatiques.

	Toutes les méta-analyses			Méta-analyses robustes		
	# études	# accord	%	# études	# accord	%
PSY						
Tests cognitifs	66	41	62,1	65	41	63,1
Imagerie	95	36	37,9	55	23	41,8
Génétiques	80	26	32,5	26	7	26,9
Autres	79	28	35,4	59	22	37,3
NEURO						
Génétiques	83	40	48,2	39	16	41
Autres	72	42	58,3	43	26	60,5
SOMA						
Génétiques	121	60	49,6	57	32	56,1
Autres	63	24	38,1	39	16	41

Les résultats ont été étudiés séparément pour toutes les méta-analyses et pour les méta-analyses robustes uniquement. Les taux de reproductibilité varient fortement en fonction du domaine biomédical et du type de facteurs analysés. Ainsi les tests cognitifs en psychiatrie ont le taux de reproductibilité le plus élevé, aux alentours de 63%. En revanche si on restreint les études d'association dans le domaine de la psychiatrie aux seuls facteurs biologiques (sans les tests cognitifs), les taux de reproductibilité des études psychiatriques sont alors bien inférieurs à ceux des autres domaines (35,4% en psychiatrie biologique contre 53,5% en neurologie et 45,7% pour les maladies somatiques). Les différences entre la psychiatrie biologique et les autres domaines sont statistiquement significatifs à $p=0,0003$ avec la neurologie et $p=0,031$ avec les autres maladies somatiques. Ainsi les taux de reproductibilité pour chacun des facteurs biologiques en psychiatrie varient entre 32 et 38%.

Pour la neurologie, les taux de reproductibilité des études initiales concernant des facteurs génétiques sont similaires à ceux des études somatiques, aux alentours de 49%. En ce qui concerne les autres facteurs, le meilleur taux de reproductibilité est atteint pour les maladies neurologiques avec un taux de reproductibilité de 58%. Par comparaison, les études initiales sur les autres facteurs ne sont reproduites qu'à 38% pour les maladies somatiques. Si l'on restreint l'analyse aux méta-analyses robustes, les résultats ne changent pas (Tableau 5). Quel que soit le domaine biomédical et le type de facteur considéré, bien que les taux de reproductibilité varient, ils n'en restent pas moins faibles avec au mieux 60% de reproductibilité.

La même analyse a été restreinte aux études initiales rapportant un effet statistiquement significatif appariées ou non aux méta-analyses robustes. Les résultats sont présentés dans tableau 6 et la figure 5.

Tableau 6 : Reproductibilité des études initiales statistiquement significatives en fonction des différents domaines et des types de facteurs de risque étudiés. Une étude est considérée comme reproduite si elle est en accord avec la méta-analyse et que le taux d'inflation du résultat de l'étude initiale n'excède pas 100%. PSY : Psychiatrie, NEURO : neurologie, SOMA : maladies somatiques

	Toutes les méta-analyses			Méta-analyses robustes		
	# études	# accord	%	# études	# accord	%
PSY						
Tests cognitifs	45	38	84,4	44	38	86,4
Imagerie	50	21	42	28	18	64,3
Génétiques	46	3	6,5	16	1	6,3
Autres	46	21	45,7	36	20	55,6
NEURO						
Génétiques	37	14	37,8	19	11	57,9
Autres	41	27	65,9	29	24	82,8
SOMA						
Génétiques	61	24	39,3	40	24	60
Autres	30	9	30	20	8	40

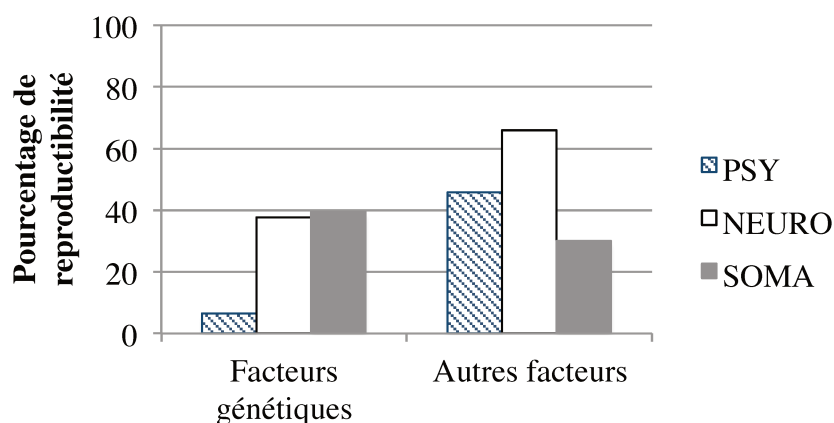


Figure 5 : Taux de reproductibilité des études initiales statistiquement significatives en fonction des différents domaines et des facteurs de risque génétiques ou des autres facteurs.

Dans le domaine de la psychiatrie, les taux de reproductibilité des études initiales sont très élevés pour les tests comportementaux ou cognitifs (plus de 84%). Ce fort taux de reproductibilité peut être expliqué par la nature tautologique de la plupart de ces tests. Par exemple, on peut effectivement s'attendre à ce qu'une plus grande impulsivité soit associée au TDAH. Pour les autres types d'association les taux de reproductibilité sont nettement plus faibles (Tableau 6 et Figure 5). Ils sont particulièrement bas pour les études génétiques en psychiatrie (6,5%). Autrement dit, ces études initiales ne sont pratiquement jamais confirmées par les méta-analyses.

2.5- Influence de la taille des échantillons et des tailles d'effet sur la reproductibilité des études initiales

Nous avons analysé l'influence du nombre médian de patients inclus dans les études initiales et de la taille d'effet « vraie », c'est à dire celle indiquée par la méta-analyse sur le taux de reproductibilité des études initiales.

Les résultats sont présentés sur un graphique à 3 dimensions pour toutes les études initiales statistiquement significatives (Figure 6).

Les résultats montrent que, dans le cas des études génétiques en neurologie et pour les maladies somatiques, les tailles d'effet réelles sont certes petites mais le grand nombre de patients permet d'obtenir des taux de reproductibilité plus élevés. Par contre, pour les études génétiques en psychiatrie, le nombre moyen de patients est trop faible pour pouvoir mesurer des tailles d'effet de petite taille. Le taux de reproductibilité est donc très médiocre.

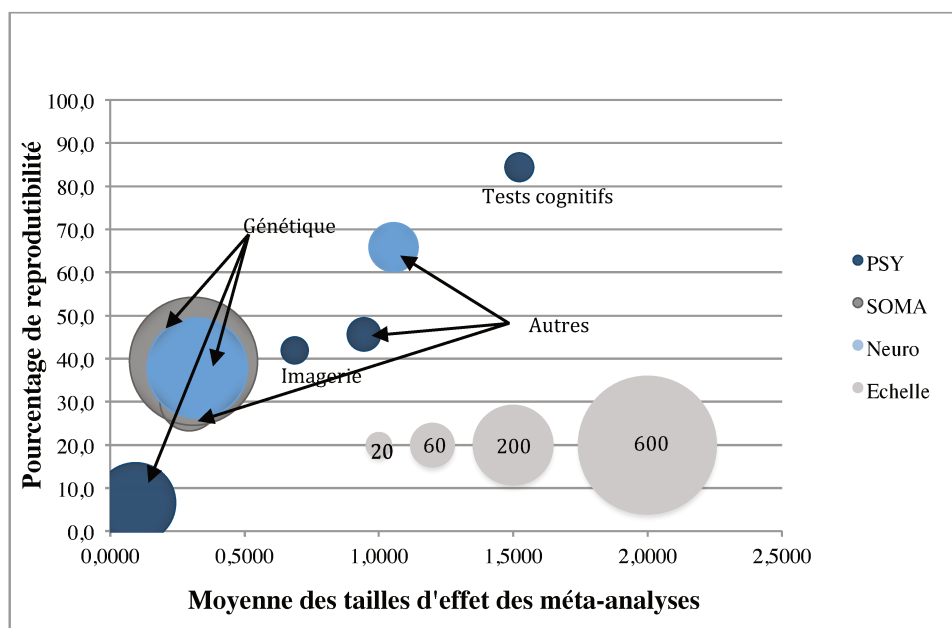


Figure 6 : Reproductibilité des 356 études initiales statistiquement significatives en fonction du type d'étude réalisé. Génétique : facteurs de risque génétiques, Autres : autres types de facteurs, Imagerie : résultats d'imagerie cérébrale (en psychiatrie), Tests cognitifs : tests cognitifs et comportementaux (en psychiatrie). PSY : psychiatrie, NEURO : neurologie, SOMA : maladies somatiques. La taille des bulles est proportionnelle au nombre de patients inclus dans l'étude initiale. L'échelle en gris donne les tailles pour 20, 60, 200 et 600 patients. La taille d'effet moyenne rapportée par les méta-analyses est considérée comme une bonne approximation de la valeur « réelle » de l'association entre le facteur et la pathologie. La reproductibilité des études tient donc compte de cet effet ainsi que du nombre de patients inclus dans l'étude initiale.

2.6- Taux de reproductibilité des études les plus larges

Nous avons vu au paragraphe 2.4.1 et sur la figure 2 que les tailles d'effet moyennes des études les plus larges, c'est à dire celles incluant le plus grand nombre de patients, sont plus souvent en accord avec celles des méta-analyses. Afin d'analyser la validité des études les plus larges, nous avons, comme pour les études initiales, déterminé leur sensibilité et leur spécificité (Tableau 7). En ce qui concerne la sensibilité, les études les plus larges sont similaires aux études initiales, elles ne parviennent pas à s'affranchir de tous les faux négatifs. En revanche, leur spécificité est bien meilleure, supérieure à 0,8 ; elles identifient mieux les faux positifs. Ces valeurs de spécificité et de sensibilité sont comparables entre les 3 domaines de recherche, et ceci même quand l'analyse est limitée aux méta-analyses robustes.

D'autre part, le taux de reproductibilité des études les plus larges est supérieur à celui des études initiales. Si l'on considère les résultats de toutes les études les plus larges, leur taux de reproductibilité - défini comme précédemment c'est à dire avec une inflation inférieure ou égale à 100%- varie entre 62,5% et 67,9% en fonction des domaines de recherche. Si l'analyse est restreinte aux études les plus larges statistiquement significatives, les taux de reproductibilité varient alors de 70,6% à 80,4%. Enfin, ce taux ne varie pas en fonction du domaine biomédical considéré.

Tableau 7 : Sensibilité et spécificité des résultats études les plus larges comparées à ceux des méta-analyses. Les intervalles de confiance sont ceux à 95%.

	Pathologies psychiatriques	Maladies neurologiques	Maladies somatiques
Méta-analyses (toutes)	320	159	184
Sensibilité	0,586 (0,518; 0,652)	0,646 (0,543; 0,738)	0,538 (0,438; 0,636)
Spécificité	0,82 (0,728; 0,887)	0,817 (0,691; 0,901)	0,85 (0,749; 0,917)
Méta-analyses robustes	205	85	95
Sensibilité	0,669 (0,594; 0,736)	0,795 (0,686; 0,875)	0,643 (0,519; 0,751)
Spécificité	0,926 (0,742; 0,987)	1 (0,598; 1)	0,875 (0,665; 0,967)

2.7- Influence du facteur d'impact du journal dans lequel l'étude initiale est publiée

Nous avons également testé si le prestige du journal (mesuré indirectement par son facteur d'impact) dans lequel l'étude initiale avait été publiée avait une influence sur le taux de reproductibilité. Dans l'ensemble, les études initiales en psychiatrie sont publiées dans des journaux avec des facteurs d'impact plus forts que pour les études en neurologie et pour les autres maladies somatiques (facteur d'impact médian en psychiatrie 7,7, en neurologie 4,8 et pour les maladies somatiques 5,6). Cependant ces différences ne sont pas statistiquement significatives.

Nous avons calculé un index d'inflation : le rapport entre la taille d'effet de l'étude initiale et celle de la méta-analyse correspondante. Nous avons alors distribué ces index d'inflation en fonction du facteur d'impact du journal dans lequel l'étude initiale avait été publiée. Après analyse des résultats par régression linéaire, nous n'avons pas trouvé de différence dans le taux de reproductibilité, quel que soit le domaine biomédical considéré ou le type de facteur analysé. Autrement dit, la publication d'une étude initiale dans un journal à fort facteur d'impact ne garantit aucunement une meilleure reproductibilité.

2.8- Influence de l'année de publication de l'étude initiale sur le taux de reproductibilité

Les 663 études initiales identifiées dans notre base de données ont été publiées entre 1958 et 2009. Seules 40 études initiales ont été publiées avant 1980. Nous avons donc formé

un groupe 1958-1979 puis 3 autres groupes correspondant aux années 1980 (1980-1989), 1990 (1990-1999) et aux années 2000 (2000-2009). Après analyse, nous n'avons pas identifié de différence du taux de reproductibilité entre ces 4 groupes aussi bien entre les domaines de recherche qu'entre les facteurs de risques étudiés. Le taux de reproductibilité des études initiales contenues dans notre base de données ne semble donc pas avoir évolué sur six décennies. Cependant, pour le premier groupe (1958-1979), le faible nombre d'observations et leur étalement sur plus de deux décennies rend la comparaison avec les trois décennies suivantes discutable.

Il est néanmoins utile de préciser une différence au niveau des études des facteurs de risque génétiques. En effet, en psychiatrie, seules deux études initiales (sur 46) ont été publiées entre 2005 et 2009. Par contre, pour la neurologie et les maladies somatiques, de nombreuses études initiales ont été publiées sur la même période (17/37 pour la neurologie et 22/61 pour les maladies somatiques). Or, pendant cette période, le nombre de patients inclus dans les études initiales a été augmenté de plus de 6 fois. Ceci permet d'expliquer en partie pourquoi les taux de reproductibilité des études génétiques sont si faibles en psychiatrie : leurs tailles d'échantillon sont en moyenne plus faibles que celles utilisées en neurologie ou pour les maladies somatiques.

2.9- Evolution du nombre moyen de patients et des tailles d'effet moyennes des études initiales

Une des explications souvent avancée pour expliquer le faible taux de reproductibilité des études initiales est le trop faible nombre d'observations. Une des mesures proposée est donc d'augmenter le nombre de patients inclus dans les études primaires.

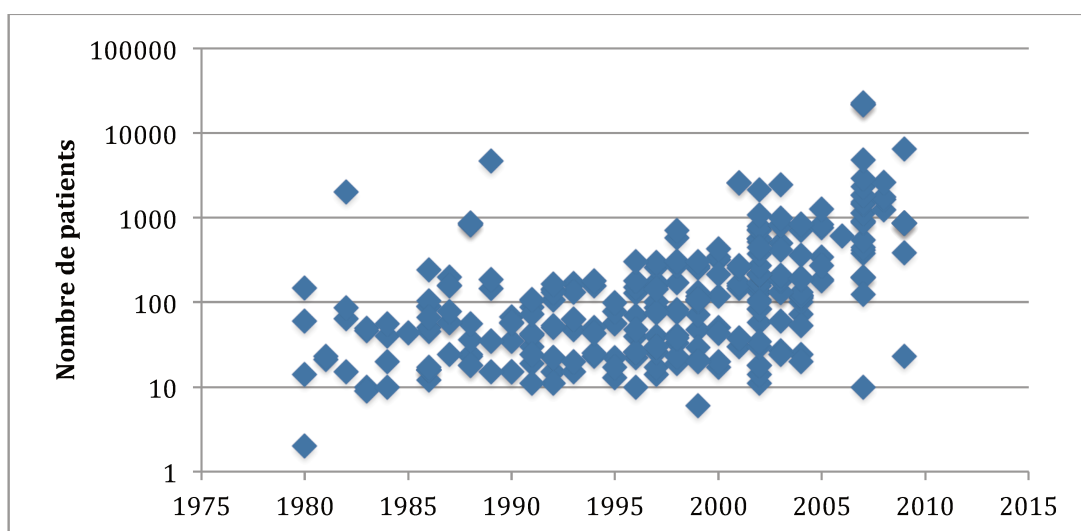


Figure 7 : répartition des tailles d'échantillons des études initiales en fonction de leur année de publication

Nous avons représenté, sur la figure 7, le nombre de patients inclus dans les 296 études initiales significatives en fonction de l'année de publication de l'étude initiale, après exclusion de celles publiées avant 1980. Le graphique montre que le nombre de patients inclus dans les études initiales augmente avec les années.

Comment expliquer alors que le taux de reproductibilité des études initiales n'ait pas progressé malgré l'augmentation du nombre moyen de patients par étude?

Nous avons alors représenté les tailles d'effet des 296 études initiales en fonction de l'année de publication (Figure 8). On remarque que les tailles d'effet observées sont de plus en plus petites. Elles sont passées d'un OR moyen aux alentours de 9 en 1980 à des OR de 1,7 en 2009. Par conséquent, même si le nombre de patients inclus dans les études initiales augmente, le taux de reproductibilité n'est pas amélioré puisque, en parallèle, les tailles d'effet rapportées comme significatives sont plus faibles.

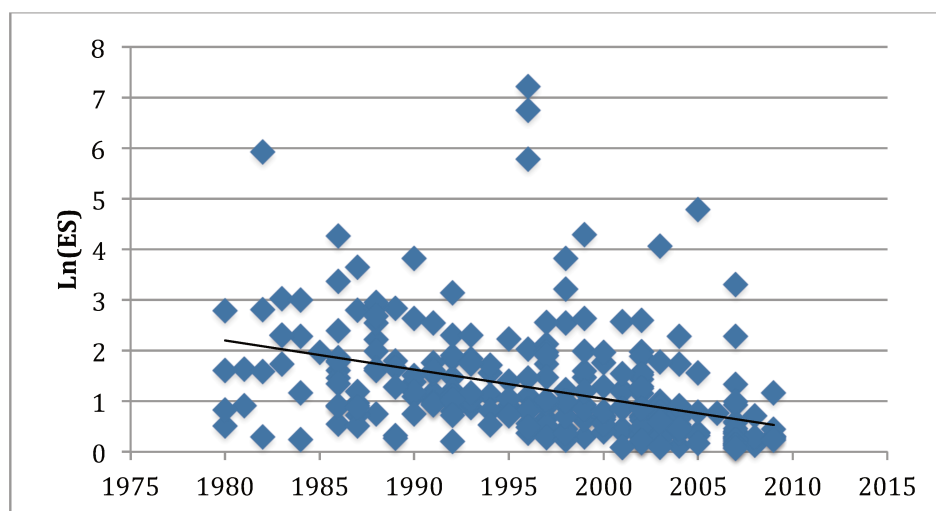


Figure 8 : répartition des tailles d'effet des études initiales en fonction de leur année de publication.

III- Apports et limites de cette étude

3.1- Les apports

Les résultats de cette étude confirment la faible reproductibilité des études initiales. Nous avons quantifié cette « crise de la reproductibilité » pour différents facteurs : (1) des tests cognitifs ou comportementaux, (2) des facteurs génétiques, (3) des résultats d'imagerie cérébrale et (4) ce que nous avons appelé les autres facteurs. Ils comprennent des mesures biochimiques ou encore des facteurs de risques environnementaux (e.g. alimentaires). Nous avons observé de fortes différences de reproductibilité en fonction du type de facteur analysé. Ainsi, les études initiales portant sur des tests cognitifs sont très souvent confirmées par les méta-analyses (84,4% pour les études initiales statistiquement significatives). A l'inverse, les

études génétiques en psychiatrie sont très rarement confirmées (seules 6,5% des études initiales significatives le sont). Dans l'ensemble, et à l'exception de ces deux cas, les taux de reproductibilité sont peu élevés entre 33 et 65% (Tableaux 5 et 6 et Figure 5). D'autre part, la base de données que nous avons créée inclut 663 études d'association dans trois grands domaines de la recherche biomédicale, la psychiatrie, la neurologie et des maladies somatiques. L'analyse de la reproductibilité des résultats des études initiales dans ces trois domaines n'a pas montré de différence majeure : les taux de reproductibilité sont médiocres dans tous les domaines, entre 40 et 60% en moyenne (Figure 4).

Nous avons choisi un seuil d'inflation $\leq 100\%$ pour évaluer les taux de reproductibilité. Bien que ce seuil soit relativement indulgent, 43% des résultats des études initiales statistiquement significatives dépassent ce seuil d'inflation. Autrement dit, les résultats de ces études initiales sont plus de deux fois plus élevés que ceux rapportés par les méta-analyses associées. Les résultats de notre étude confirment donc la dévaluation (« *decline effect* ») des résultats des études initiales (Brembs et al. 2013; Fanelli et al. 2017). Pour chacune des 12 pathologies, les résultats moyens des études initiales sont tous largement au dessus de la droite d'équivalence qui représente l'accord entre les tailles d'effet des études initiales et celles des méta-analyses. Or, il n'y a pas de raison statistique pour expliquer que les résultats des études initiales surestiment plus qu'ils ne sous-estiment la taille d'effet.

Des auteurs ont suggéré que cette dévaluation pourrait être expliquée par des tailles d'échantillon trop petites. Elles exacerberaient les problèmes de reproductibilité et favoriseraient la surestimation des tailles d'effet (Button et al. 2013; Ioannidis 2005, 2008). D'autres études quantitatives (Ioannidis et al. 2003; Pereira et al. 2012; Trikalinos et al. 2004), ainsi que l'étude décrite ici sont en accord avec ces hypothèses. Notre étude montre néanmoins que la relation entre la taille d'effet et la taille des échantillons est relative. En effet, si les tailles d'effet sont « réellement » élevées, le nombre de patients sur lequel la mesure est réalisée ne nécessite pas d'être très grand. C'est le cas pour les tests cognitifs. En revanche si la taille d'effet recherchée est petite, comme c'est le cas par exemple pour les facteurs de risque génétiques, le résultat ne sera robuste que s'il est réalisé sur un grand échantillon. Ainsi, le cas des études génétiques en psychiatrie le confirme : leur taux de reproductibilité est très faible même si, proportionnellement aux études des autres facteurs de risque, elles comprennent un plus grand nombre de patients. En effet, le nombre moyen de patients inclus dans les études génétiques est de 200 alors que les études concernant les autres facteurs n'en comprennent qu'une trentaine. Ce nombre moyen reste cependant bien inférieur à celui inclus pour les maladies neurologiques (320) ou les maladies somatiques (520). Mais, dans le cas des études génétiques en psychiatrie, les tailles d'effet « réelles », données par les

méta-analyses, sont très faibles. Enfin, les analyses des résultats des études les plus larges confirment l'importance de la taille des échantillons dans la reproductibilité des études. En effet, les résultats de ces études sont plus souvent en accord avec ceux des méta-analyses (Figure 2) et leur taux de reproductibilité est plus élevé (Tableau 7).

Nous avons également analysé l'influence du facteur d'impact du journal scientifique sur la reproductibilité des études initiales. En effet, les scientifiques cherchent à publier leurs résultats dans des journaux prestigieux (Brembs et al. 2013). C'est un processus difficile et très sélectif. Or, nous n'avons pas observé de différence du taux de reproductibilité quel que soit le facteur d'impact du journal. Par conséquent, les résultats publiés dans des journaux à fort facteur d'impact ne sont ni plus ni moins fiables que des résultats publiés dans des journaux moins prestigieux. Cette conclusion est en accord avec deux autres études (Munafò et al. 2009; Prinz et al. 2011).

On parle de la « crise de la reproductibilité » depuis maintenant une vingtaine d'années et des observations quantitatives ont été faites depuis les années 2000 par l'équipe de John Ioannidis notamment (Ioannidis et al. 2001b; Ioannidis 2005b). Par contre, l'évolution temporelle du taux de reproductibilité n'avait, à notre connaissance, jamais été étudiée. Notre étude montre que ce taux n'a pas évolué pendant trois décennies (1980-2009). Après la publication de notre étude dans PLoS One, nous avons entrepris d'expliquer cette stabilité en mesurant l'évolution temporelle de la taille d'échantillon et de la taille d'effet. En effet, de nombreux auteurs suggèrent que les problèmes de reproductibilité sont en partie dus aux études réalisées sur de trop petits échantillons (Button et al. 2013; Ioannidis et al. 2003). Ce constat aurait dû pousser les chercheurs à augmenter le nombre d'observations afin d'augmenter la reproductibilité de leurs résultats. Entre 1980 et 2009, nous avons effectivement observé une augmentation importante de la taille des échantillons, sans que celle-ci se traduise par une amélioration de la fiabilité. En effet, pendant la même période, la taille des effets considérés comme significatifs n'a cessé de baisser. Ainsi, comme nous l'avons déjà souligné, plus la taille d'effet réelle est petite, ou autrement dit plus l'association entre deux facteurs est ténue, plus la taille des échantillons devra être importante pour être en mesure de la détecter. Pour palier à ce problème, les études d'associations entre des facteurs génétiques et des pathologies sont maintenant réalisées sur de multiples génotypes et incluent tout le génome. Elles sont, de ce fait, réalisées en collaboration, sur de nombreuses populations pour pouvoir être en mesure de détecter ces éventuelles associations (Ioannidis 2013).

3.2- Les limites

Nous avons choisi d'utiliser les résultats de méta-analyses pour quantifier la reproductibilité des études initiales. Les méta-analyses sont dès lors considérées comme donnant la meilleure estimation de la force d'une association entre des facteurs de risque et des pathologies. Nous avons donc choisi des méta-analyses robustes : elles compilent les données d'au moins sept populations différentes issues d'au moins quatre publications indépendantes. En effet, une étude a comparé les résultats de méta-analyses publiées à cinq ans d'intervalle sur la même association (Pereira et al. 2011). Les auteurs ont ainsi montré que plus le nombre de populations inclus dans la méta-analyse était grand et moins la taille d'effet indiquée par la méta-analyse était modifiée. Cependant, nous ne pouvons pas exclure que les méta-analyses que nous avons sélectionnées présentent elles-aussi des effets exagérés ou mêmes non reproductibles. Si les méta-analyses restent la meilleure estimation de la mesure de la force d'une association, il n'en reste pas moins qu'elles sont réalisées à partir d'études publiées. Le fonctionnement actuel de la recherche favorise les études positives et les méta-analyses sont souvent affectées par ce biais de publication, en particulier celles qui incluent un petit nombre d'études primaires distinctes. D'autre part, la qualité méthodologique des études primaires incluses dans les méta-analyses peut être variable (Sterne et al. 2001). Plusieurs auteurs ont proposé des tests pour identifier ces biais et essayer d'en tenir compte (Egger et al. 1997; Ioannidis 2011d; Thompson et al. 2011). Cependant ces méthodes nécessitent bien souvent de modéliser indépendamment chaque source de biais et sont donc difficiles à utiliser dans le cadre de cette étude qui couvre un grand nombre de méta-analyses. Pour évaluer ces biais, nous avons donc, dans un deuxième temps, limité notre analyse à une sous-population de méta-analyses robustes. Les critères de robustesse ont été définis en accord avec la littérature (Ioannidis 2008b; Pereira et al. 2011).

Le choix des pathologies étudiées ici est partiellement arbitraire. En effet, l'utilisation de méta-analyses pour mesurer la reproductibilité des études initiales présente des limites. La principale est le nombre de méta-analyses publiées sur un sujet donné. Par conséquent, les méta-analyses portant sur des maladies rares sont peu nombreuses. Ensuite, nous n'avons pas inclus dans notre analyse certaines maladies neurologiques telles que l'accident vasculaire cérébral ou les traumatismes crâniens puisqu'elles sont aussi des maladies cardiovasculaires ou que leur survenue est accidentelle. Enfin, nous avons choisi, pour chacun des trois domaines, des pathologies dont les prévalences dans la population étaient similaires.

Comme nous l'avons mentionné dans l'introduction de ce chapitre, la trop petite taille d'échantillons n'explique qu'en partie la faible reproductibilité des études primaires. D'autres facteurs sont impliqués, comme la publication préférentielle des résultats positifs ou encore

des problèmes méthodologiques (changement de l'hypothèse à l'issue des résultats, sélection des données et de l'analyse statistique pour obtenir un résultat significatif) éventuellement exacerbés par des conflits d'intérêt (Fanelli et al. 2017; Ioannidis 2005; Munafo et al. 2017). Cette étude ne permet pas de mettre en évidence ces autres facteurs.

La reproductibilité des résultats de la recherche biomédicale est ici limitée à l'analyse de la reproductibilité des études initiales et des études les plus larges. En effet, nous ne pouvions pas examiner la reproductibilité de toutes les études primaires¹⁷ (10154 résultats) incluses dans notre base de données. Néanmoins, la quantification de la reproductibilité des études initiales est essentielle. En effet, ces études influencent et conditionnent l'étendue des recherches sur un sujet (Trikalinos et al. 2004b) et sont plus volontiers sélectionnées par les médias (Gonon et al. 2012 ; voir aussi le chapitre suivant).

Enfin, nous avons quantifié la reproductibilité des études initiales uniquement lorsque celles-ci étaient incluses dans des méta-analyses contenant au moins sept populations différentes. Nous n'avons pas analysé la reproductibilité des études initiales incluses dans des méta-analyses moins étendues ou n'ayant pas fait l'objet de méta-analyses. Notre analyse est donc limitée à une petite fraction des études initiales de la recherche biomédicale. En effet, dans un échantillon de 259 études empiriques en recherche biomédicale seules 16 d'entre elles (6,2%) ont été incluses dans des méta-analyses (Iqbal et al. 2016).

IV- Conclusion

La méthode scientifique part de résultats prometteurs mais incertains pour arriver à un consensus après réplification de ces résultats par d'autres scientifiques. Les chercheurs formulent une hypothèse, la teste par des expériences et rédigent alors un projet de publication qui est soumis à l'examen par les pairs (« *reviewers* »). Ceux-ci contrôlent les méthodes utilisées, la cohérence interne des résultats, leur accord avec les conclusions et leur discussion à la lumière d'un examen non biaisé de la littérature. Puis ils acceptent ou non la publication. Après publication, les scientifiques du domaine concerné pourront alors valider ou non ces résultats en reproduisant l'étude. Il est donc normal que tous les résultats ne soient pas reproductibles, c'est le mode de fonctionnement de la recherche, par tâtonnements et vérification des résultats.

¹⁷ Les résultats primaires correspondent à chaque ensemble de données relatives à l'association analysée. Par exemple, pour un polymorphisme génétique une même publication peut comprendre l'analyse de plusieurs polymorphismes qui seront tous des résultats indépendants (taille d'effet, taille de l'échantillon, intervalle de confiance)

Classiquement, la combinaison de cet examen par les « *reviewers* » et la réplication par des chercheurs concurrents représente le socle qui fonde la validité des résultats scientifiques. Cependant, comme le confirme notre étude, les résultats des études initiales sont peu reproductibles et les tailles d'effet qu'elles présentent sont largement supérieures à celles des méta-analyses. Les résultats des études initiales sont donc bien souvent exagérés. Ces observations, réalisées dans le domaine de la recherche biomédicale, sont en accord avec les inquiétudes exprimées dans les journaux scientifiques : plus de la moitié des résultats publiés seraient faux et, plus grave encore, ce serait le mode de fonctionnement de la recherche qui serait en cause (Horton 2015; Smaldino et al. 2016) . En effet, d'autres études ont montré que la plupart des études initiales étaient contredites par les études ultérieures dans d'autres domaines comme la psychologie (Open Science Collaboration 2015). De nombreuses raisons, que nous avons déjà citées, sont avancées. Insistons encore sur le fait que les rares cas de fraude caractérisée ne contribuent que pour une part très mineure à la « crise de la reproductibilité » telle qu'elle est décrite dans la littérature biomédicale. Les causes principales semblent plutôt être internes au mode de fonctionnement actuel de la recherche. En effet, le système de publication favorise la présentation de résultats extraordinaires, positifs et décourage la réplication (Lawrence 2003; Young et al. 2008). Ainsi, il incite indirectement à la poursuite de pratiques de recherche discutables pour optimiser les résultats des recherches et donc leur publication (Fanelli 2010). D'autre part, les canons statistiques utilisés (tailles d'échantillon, puissance statistique et probabilité à $p < 0,05$) sont aussi à blâmer (Button 2016; Horton 2015). Cet excès d'études initiales positives exagérées est aussi le résultat de biais cognitifs : la tendance à favoriser des résultats qui vont dans le sens de l'hypothèse et la tendance à considérer qu'un résultat était prédictible une fois qu'il a été observé (Munafo et al. 2017).

De nombreuses suggestions ont été faites pour palier à ces différents problèmes. Tout d'abord améliorer la description des résultats pour permettre une reproduction et une analyse des résultats plus faciles (Glasziou et al. 2014). Des listes d'instructions ont ainsi été publiées par EQUATOR (<http://www.equator-network.org>) pour différents domaines : les études précliniques sur les animaux (ARRIVE), les méta-analyses (PRISMA), les études observationnelles en épidémiologie (STROBE) ou encore pour les essais cliniques randomisés (CONSORT). D'autres suggestions impliquent par exemple l'accès à toutes les données de recherche, la promotion de méthodes d'analyse rigoureuses ou encore la pré-inscription et la validation du protocole de recherche (Ioannidis et al. 2014b; Macleod et al. 2014; Nosek et al. 2012). Enfin les initiatives telles que *PubPeer*, ou *PubMed Commons* sont encouragées pour permettre aux articles scientifiques d'être discutés et commentés après

publication (Munafo et al. 2017). Cependant, de telles initiatives sont aussi critiquées : chaque incitation positive a un revers (Bissell 2013; Lewandowsky et al. 2016). Par exemple, l'accès aux données de recherche sans restriction peut avoir des conséquences éthiques néfastes. Ainsi, les patients qui ont donné leur accord pour participer à une étude ne s'attendent pas à voir leurs données utilisées par d'autres équipes de recherche. Aussi, toutes ces pré-inscriptions, pré-publications, listes d'instructions alourdissent encore les responsabilités administratives des chercheurs et découragent la poursuite de recherches créatives.

D'autres auteurs s'inquiètent des conséquences de cette mauvaise reproductibilité des résultats de la recherche biomédicale en raison de son coût économique (Freedman et al. 2015; Gan et al. 2012; Macleod et al. 2014). Par exemple, sur 253 essais cliniques randomisés évaluant des thérapies anti-cancéreuses, 158 ont été négatifs en raison d'une surestimation de l'efficacité du traitement lors de la mise en place de l'essai clinique (Gan et al. 2012). Les auteurs estiment le coût de ces 158 échecs à 10 milliards de dollars. D'autre part, les auteurs soulignent aussi que 100275 patients ont été inutilement enrôlés dans ces essais cliniques - qui se sont avérés négatifs - et que 5287 patients ont été exposés à des traitements dont la nocivité a été prouvée. Cette médiocre reproductibilité a également des conséquences sur les acteurs de la recherche eux-mêmes. Ainsi, des enquêtes ont été réalisées auprès de chercheurs en sciences biomédicales (Siebert et al. 2015; Tijdink et al. 2016). Elles révèlent des sentiments négatifs et des vues pessimistes voire cyniques sur la qualité et la validité de la recherche scientifique. Enfin, cette « crise de la reproductibilité » pourrait aussi avoir des conséquences négatives sur la confiance du public. Ces problèmes de reproductibilité sont en effet relayés dans les journaux grand public (Lehrer 2010; The Economist 2013a) ou des ouvrages de vulgarisation (Harris 2017).

Enfin, ce manque de reproductibilité est problématique pour la pratique scientifique elle-même¹⁸. En effet, les exagérations des résultats d'études initiales sont vraisemblablement dues aux politiques d'évaluation de la recherche qui utilisent des indicateurs numériques. Pour obtenir un poste ou des financements, le nombre et le prestige des journaux scientifiques dans lesquels les articles sont publiés sont cruciaux (Lawrence 2003, 2007). Les journaux prestigieux favorisent la publication de résultats spectaculaires, susceptibles d'attirer l'attention des médias (Bucchi 2015; Franzen 2012). Or, ces journaux prestigieux sont régulièrement utilisés comme sources par les journalistes (Hansen 1994; Stryker 2002; Weitkamp 2003). Comme nous l'avons confirmé dans cette étude, les résultats publiés dans ces journaux ne sont pas plus valides que ceux publiés dans d'autres journaux à

¹⁸ Ceci sera plus largement discuté dans le dernier chapitre (Chapitre V)

comité de lecture mais moins prestigieux. Cependant, à notre connaissance, peu d'études ont été conduites dans le but de déterminer quels types de résultats scientifiques étaient choisis et présentés dans les médias. En effet, des études ont analysé l'influence des communiqués de presse sur la couverture médiatique d'articles scientifiques (de Semir et al. 1998) ou encore quels articles étaient sélectionnés dans les journaux biomédicaux (Bartlett et al. 2002). Par contre, la couverture médiatique des résultats d'études initiales n'a pas été analysée si ce n'est dans le cas du TDAH (Gonon et al. 2012). Ceci nous semble un point important à étudier. En effet, si la reproductibilité des études initiales est médiocre, il est vraisemblable que l'information présentée au public dans les médias soit exagérée voire fausse. Dans le prochain chapitre (chapitre III), nous décrivons la couverture médiatique des études contenues dans la base de données que nous avons décrite ici et dont nous connaissons la validité. Nous avons en particulier analysé les facteurs qui influençaient la sélection des études scientifiques par la presse écrite ainsi que la façon dont la presse traitait de l'incertitude.

CHAPITRE III

Incertitude et médias - Quantification et analyse de la couverture médiatique des résultats de la recherche biomédicale

La principale source d'information du public concernant la recherche biomédicale vient des médias (Nelkin 1995). Dans une enquête récente commanditée par la direction générale de la recherche de la Commission européenne, 41% des européens interrogés ont indiqué que la presse écrite était leur source principale d'information sur les résultats de la recherche juste derrière la télévision (68%). D'autre part, 62% des personnes interrogées indiquent que la médecine est le domaine de recherche qui les intéresse le plus (Eurobaromètre 2007).

Il est donc important de comprendre comment les journalistes choisissent et présentent les résultats scientifiques qu'ils vont couvrir. Plusieurs études se sont attachées à définir les critères qui attiraient et donc orientaient le choix des journalistes. Les journalistes favorisent les résultats scientifiques qui ont une implication directe pour le lecteur, tels que des études scientifiques impliquant le mode de vie (« *lifestyle* »), l'exposition à des maladies ou à des produits toxiques (Hansen 1994; National Academies of Sciences 2017; Stryker 2002; Weitkamp 2003). Les journalistes s'intéressent aussi aux sujets qui sont controversés (National Academies of Sciences 2017) ou encore porteurs de scénarios dramatiques (Wilkins et al. 1987). Enfin, les nouveautés ou découvertes scientifiques sont aussi prisées (Logan et al. 2000).

D'autres études se sont intéressées aux sources choisies par les journalistes. Elles ont montré que les journalistes s'appuyaient sur les communiqués de presse publiés par les institutions scientifiques et les journaux scientifiques (de Semir et al. 1998; van Trigt et al. 1994). Les sommaires et éditoriaux des grands journaux scientifiques sont aussi des sources fréquemment utilisées par les journalistes (Entwistle 1995; Suleski et al. 2010). En effet, les journalistes considèrent les journaux scientifiques prestigieux, tels que *Science*, *Nature* ou encore *The Lancet*, comme des sources crédibles et fiables d'informations (Hansen 1994). Ils sont donc tributaires de la qualité et de la fiabilité des publications scientifiques (Wilkie 1996).

On peut cependant s'interroger sur la validité des publications scientifiques. En effet, nous venons de décrire dans le chapitre précédent la « crise de la reproductibilité » qui entache la recherche biomédicale. Elle touche tous les domaines : que ce soit les études précliniques (Freedman et al. 2015; Prinz et al. 2011), les essais cliniques (Gan et al. 2012; Ioannidis et al. 2001; Tajika et al. 2015) ou les études d'association entre des facteurs de risques ou des marqueurs biologiques et des pathologies (Dumas-Mallet et al. 2016; Ioannidis et al.

2011c; Trikalinos et al. 2004b). Nous avons confirmé que les résultats des études initiales étaient souvent exagérés et contredits par les méta-analyses (Dumas-Mallet et al. 2016). Nous avons aussi montré que le taux de reproductibilité ne variait pas avec le facteur d'impact du journal. Or, ces études initiales décrivent bien souvent des résultats spectaculaires bien qu'incertains. Leur réplification est donc indispensable pour assurer leur validité. Malheureusement, les éditeurs de revues prestigieuses favorisent la publication de résultats « révolutionnaires » susceptibles d'attirer l'attention des médias (Bucchi 2015). Certains chercheurs en sont conscients et tendent à exagérer l'intérêt de leurs travaux pour être publiés dans ces revues (Gonon et al. 2011; Lawrence 2003). Par conséquent, la qualité et la fiabilité des sources d'information privilégiées par les médias ne sont certainement pas optimales.

Cependant, à notre connaissance, la couverture préférentielle des études initiales n'a pas été étudiée, si ce n'est dans le cas du trouble déficitaire de l'attention avec hyperactivité (Gonon et al. 2012). Dans cette étude, les auteurs ont analysé la reproductibilité de dix études scientifiques relatives au TDAH qui avaient fait l'objet de nombreux articles de presse dans les années 1990. Parmi ces dix études, sept étaient des études initiales. Lorsque les résultats de ces sept études ont été comparés à ceux des études ultérieures sur le même sujet, quatre ont été invalidées et les trois autres présentaient des tailles d'effet largement surestimées. Cette étude n'a cependant examiné qu'un petit échantillon de publications scientifiques et une seule pathologie. Nous avons donc utilisé la base de données décrite dans le chapitre précédent pour analyser la couverture médiatique des publications scientifiques. En effet, cette base de données rassemble les résultats de 663 méta-analyses et de leurs études primaires associées. Elle présente plusieurs avantages. Premièrement, elle comporte un grand nombre d'études d'association entre de facteurs de risque ou de marqueurs biologiques et des pathologies. Ce type d'étude suscite souvent l'intérêt des journalistes et du public. Cette base de données représente donc un matériel de choix pour analyser la couverture médiatique des études scientifiques. Deuxièmement, toutes les études primaires sont appariées à des méta-analyses. Comme le résultat d'une méta-analyse est considéré comme une bonne estimation de la taille d'effet « réelle » d'une association, cette base de données permet d'évaluer la validité de chaque étude primaire et notamment de celles couvertes par les médias.

Dans ce chapitre nous décrivons les résultats en deux parties. Dans la première, nous avons quantifié le pourcentage d'études couvertes par les médias et avons analysé leur reproductibilité. Nous avons testé l'influence de différents facteurs, comme par exemple le facteur d'impact du journal dans lequel l'étude scientifique a été publiée. Enfin, nous avons examiné si des études scientifiques infirmant les résultats d'études précédemment couvertes étaient relayées par la presse. Dans la deuxième partie, nous avons réalisé des analyses de

contenu sur une sous-population d'articles de presse couvrant 40 études scientifiques initiales. Nous avons examiné en particulier la façon dont les journalistes décrivaient les résultats de la recherche et notamment leur caractère incertain. Une présentation plus détaillée sera proposée au début de la deuxième partie.

La détermination de la quantité et du type d'études scientifiques (i.e. étude initiale/ultérieure, mode de vie) couvertes par la presse devrait nous éclairer sur les critères susceptibles d'intéresser les journalistes. D'autre part, en analysant le contenu des articles de presse couvrant des études initiales, nous pourrions analyser la façon dont les journalistes tiennent compte ou non de l'incertitude inhérente aux résultats de ces études.

I- Couverture médiatique des résultats de la recherche biomédicale

Nous avons utilisé ici la base de données décrite dans le chapitre précédent pour étudier trois questions. Premièrement, nous avons quantifié et caractérisé la couverture médiatique des études primaires et des méta-analyses contenues dans la base de données. Deuxièmement, nous avons examiné la validité des études scientifiques couvertes par les médias : les résultats de ces études ont-ils été confirmés, atténués ou réfutés par les méta-analyses correspondantes ? Nous avons testé l'influence de différents paramètres sur la reproductibilité de ces études : le type d'étude couverte (étude initiale ou étude ultérieure, résultats positifs ou résultats négatifs), le facteur d'impact du journal dans lequel l'étude a été publiée, le domaine de recherche (neurologie, psychiatrie ou maladies somatiques) ou encore le type d'association analysé (i.e. lié ou non au mode de vie). Troisièmement, lorsqu'une étude couverte par les médias avait été par la suite contredite, nous avons cherché si les études la contredisant avaient elles aussi été relayées par la presse. Ces résultats ont été publiés dans la revue PLoS One en février 2017 (Dumas-Mallet et al. 2017). L'article est présenté en annexe 2.

1.1- Méthodologie

1.1.1- Sélection des études scientifiques

Nous avons utilisé la base de données décrite dans le chapitre précédent. Elle compile les résultats d'études d'association entre des facteurs de risque ou des marqueurs biologiques et des pathologies appartenant à trois domaines de recherche : la psychiatrie (Autisme, Trouble Déficitaire de l'Attention avec Hyperactivité (TDAH), schizophrénie et dépression) la neurologie (maladies d'Alzheimer et de Parkinson, épilepsie, sclérose en plaques) et quatre maladies somatiques (cancer du sein, psoriasis, glaucome et polyarthrite rhumatoïde). Cette

base de données contient les résultats de 663 méta-analyses. Au cours d'analyses préliminaires, nous avons constaté que les études concernant des tests cognitifs ou comportementaux n'étaient pas couvertes dans les médias. Nous avons donc restreint l'analyse de la couverture médiatique aux 595 méta-analyses rapportant des résultats d'association entre des facteurs génétiques, des résultats d'imagerie cérébrale ou des facteurs « autres ». Ces « autres » facteurs couvrent des facteurs épidémiologiques, biochimiques ou encore biophysiques. Des auteurs ont montré que les études impliquant le mode de vie des patients attiraient souvent l'intérêt des médias (Bartlett et al. 2002; Stryker 2002). Nous avons donc modifié la classification des études d'association de notre base de données et les avons réparties en deux catégories : « mode de vie » que nous avons appelée S et « non S ». Ainsi, la catégorie « S » regroupe des études de facteurs de risque contre lesquels le sujet peut agir, comme par exemple fumer ou manger de la viande rouge. L'autre catégorie, « non S » regroupe des facteurs de risque contre lesquels le sujet ne peut se prémunir, comme des facteurs génétiques par exemple.

Enfin, nous avons quantifié la couverture médiatique des résultats de 595 méta-analyses publiées dans 306 articles scientifiques différents et de leurs études primaires publiées dans 4775 articles scientifiques distincts. Nous avons distingué les études initiales (les premières publiées sur le sujet) des études ultérieures. Nous avons utilisé la base de données *Dow Jones Factiva* (<https://www.dowjones.com/products/factiva/>) pour identifier quelles études scientifiques avaient été couvertes par la presse écrite. Cette base de données regroupe les fils d'actualité de plus de 33000 sources internationales publiés après le 2 janvier 1981. Nous avons donc exclu les études publiées avant 1981 de notre base de données. Par conséquent, nous avons quantifié la couverture médiatique de 5029 publications scientifiques ainsi distribuées : 405 études initiales, 4218 études ultérieures et 306 méta-analyses.

1.1.2- Influence du facteur d'impact des journaux scientifiques dans lequel les études ont été publiées

Nous avons déterminé le facteur d'impact¹⁹ des journaux dans lequel les 5029 études ont été publiées. Nous avons utilisé le *Journal Citation Reports* du *Web of Science* (Thompson Reuters) et sélectionné le facteur d'impact de 2012. Dans leur analyse de la couverture médiatique du TDAH, les auteurs n'avaient pas identifié d'articles scientifiques couverts par la presse et publiés dans des journaux avec des facteurs d'impact inférieurs à 5 (Gonon et al. 2012). Nous avons donc limité initialement notre recherche sur la base de

¹⁹ En anglais *Impact Factor* (IF)

données *Dow Jones Factiva* aux articles scientifiques publiés dans des journaux à facteur d'impact supérieur ou égal à 5. Nous avons constaté que pour toutes les pathologies analysées, à l'exception du cancer du sein, les articles scientifiques couverts par les médias étaient publiés dans des journaux avec un facteur d'impact au moins égal à 6,5. Les 220 articles scientifiques publiés dans des journaux scientifiques avec des facteurs d'impact entre 5 et 6,5 n'étaient pas couverts. Il semble donc peu vraisemblable que les 2342 études scientifiques publiées dans des journaux à facteur d'impact inférieur à 5 aient attiré l'attention des journalistes. Nous n'avons donc pas recherché l'éventuelle couverture médiatique de ces articles scientifiques. Par contre, dans le cas du cancer du sein, nous avons identifié 64 articles scientifiques couverts par la presse, dont 14 avaient été publiés dans des journaux avec des facteurs d'impact compris entre 5 et 6,5. Nous avons donc étendu la recherche aux 340 articles scientifiques publiés dans des journaux scientifiques avec des facteurs d'impact compris entre 4 et 5. Nous n'en avons identifié aucun qui soit couvert par la presse. Par conséquent, nous n'avons pas non plus analysé les 339 articles relatifs au cancer du sein et publiés dans des journaux avec des facteurs d'impact inférieurs à 4.

Nous avons distingué 3 classes de facteur d'impact. La première classe comporte les facteurs d'impact supérieurs ou égaux à 30. Elle est restreinte à cinq journaux scientifiques : *New England Journal of Medicine (NEJM)*, *The Lancet*, *Journal of the American Medical Association (JAMA)*, *Nature* et *Science*. Nous avons défini la deuxième classe de manière à ce que le nombre d'articles couverts par la presse soit équivalent entre les deux dernières classes. La deuxième classe contient donc les journaux avec des facteurs d'impact entre 10 et 30 et la troisième classe, les journaux avec des facteurs d'impact inférieurs ou égaux à 10.

1.1.3- Identification des articles scientifiques couverts par la presse

Nous avons utilisé la base de données *Dow Jones Factiva*. Nous avons utilisé les mêmes mots clés pour toutes les pathologies : *study OR research OR scientist*. Nous avons ajouté d'autres mots clés spécifiques à la pathologie et au type d'étude (par exemple *genetic*, *milk* ou *birth weight*). Lorsque le nombre d'items proposé était trop grand, nous avons ajouté d'autres mots clés, comme les noms des auteurs ou de l'université.

La recherche était restreinte à un mois après la publication dans le journal scientifique et aux articles de la presse généraliste publiés en anglais. Nous avons exclu les articles publiés dans les journaux spécialisés comme *Pharma Business Week* et par les agences de presse comme *Associated Press* ou *Reuters*.

Nous avons parfois identifié des articles de presse qui couvraient une étude primaire de notre base de données, mais qui ne mentionnaient pas les résultats inclus dans la méta-

analyse correspondante. Par exemple, nous avons identifié des articles de presse qui couvraient une étude scientifique portant sur un facteur de risque pour le cancer du colon et celui du sein. Les articles de presse qui ne parlaient que du cancer du colon ont été exclus. Cette recherche a été réalisée par un des auteurs (EDM) et une analyse aléatoire de la couverture médiatique de 20% des études scientifiques a été réalisée par un autre auteur (FG), afin de s'assurer de la fiabilité de la recherche. Toutes les études scientifiques couvertes par la presse ont été correctement identifiées par les deux auteurs. Les seules différences concernaient le nombre d'articles de presse et ont été résolus après discussion.

1.1.4- Reproductibilité des études scientifiques couvertes par la presse

La reproductibilité des études couvertes par la presse a été analysée en comparant les résultats de ces études primaires à ceux des méta-analyses correspondantes comme cela a été décrit dans le chapitre II. Comme précédemment, nous avons considéré qu'une étude était reproduite lorsque (1) cette étude et la méta-analyse correspondante rapportaient une absence d'effet ou si (2) toutes deux rapportaient un effet significatif dans la même direction à condition que la taille d'effet de l'étude soit inférieure au double de celle de la méta-analyse (inflation inférieure à 100%). Certaines publications scientifiques rapportaient les résultats de plusieurs associations comme c'est le cas par exemple pour des polymorphismes génétiques. Dans ces cas, les articles de presse ne différenciaient en général pas les polymorphismes et parlaient d'un « nouveau gène ». Nous avons alors choisi le résultat le plus significatif décrit dans l'étude pour le comparer à celui de la méta-analyse. Par conséquent, la reproductibilité de chaque étude scientifique couverte par la presse a été analysée dans les conditions les plus favorables à l'étude primaire considérée.

1.2- Résultats

1.2.1- Etudes scientifiques couvertes par la presse

Nous avons identifié 156 études primaires couvertes par la presse, soit 3,3% des études incluses dans la base de données. Ce pourcentage est élevé. En effet, une étude analysant la proportion d'articles scientifiques couverts par le journal *Times* ou le programme télévisuel *NBC News* a montré que seuls 0,034% des articles scientifiques publiés étaient couverts par les médias (Suleski et al. 2010). Ces 156 études ont été couvertes par 1475 articles de presse différents. Elles concernent les 3 domaines de recherche biomédicale mais pas toutes les pathologies. En effet, aucune des études contenues dans la base de données concernant l'épilepsie et le psoriasis n'ont attiré l'attention de la presse. Cinq articles

scientifiques rapportant les résultats de méta-analyses ont également été couverts dans 86 articles de presse, soit 1,63% des articles scientifiques rapportant des méta-analyses.

Chacune des 161 études couvertes par la presse a fait l'objet de 9 articles de presse en moyenne ; certaines études attirant seulement un article de presse tandis que d'autres étaient couvertes par 50 articles de presse différents. Le nombre d'articles de presse est corrélé positivement au facteur d'impact de la revue scientifique qui a publié l'étude (régression linéaire $R^2 = 0,107$, $p < 0,0001$). Autrement dit, les études publiées dans des revues prestigieuses reçoivent une couverture médiatique plus importante.

1.2.2- Influence du domaine biomédical et du type d'association analysé

Nous avons analysé l'influence du domaine biomédical (psychiatrie, neurologie ou maladies somatiques) et du type d'association (« S » et « non S ») sur la couverture médiatique des études scientifiques. Les pourcentages varient entre 1,74% et 10,5% (Tableau 1). Les études de la catégorie « S » attirent toujours plus l'attention de la presse, quel que soit le domaine biomédical (comparer les colonnes 3 et 4 et les lignes 7 à 9 du Tableau 1). Les résultats sont similaires pour les méta-analyses ; la proportion de méta-analyses « S » couvertes est plus grande (5,4% des méta-analyses « S » contre 1,1% des méta-analyses « non S »). Il n'y a pas de différence majeure dans la proportion d'études couvertes en fonction des domaines biomédicaux (2% d'articles couverts pour la psychiatrie, 3,2% pour la neurologie et 5% pour les maladies somatiques).

Tableau 1 : Nombre d'articles primaires inclus dans les méta-analyses répartis en domaines de recherche (PSY : Psychiatrie, NEURO : neurologie et SOMA : maladies somatiques) et en type d'association (« S », « non S »). Nombre d'articles scientifiques couverts par la presse. Les valeurs entre parenthèses indiquent les pourcentages d'études couvertes.

Articles Primaires	Tous	« S »	« Non S »
PSY	1905	63	1842
NEURO	1279	158	1121
SOMA	1539	418	1121
Tous les domaines	4723	639	4084
<i>Articles couverts</i>			
PSY	38 (1,99)	6 (9,52)	32 (1,74)
NEURO	41 (3,21)	13 (8,23)	28 (2,50)
SOMA	77 (5,00)	44 (10,53)	33 (2,94)
Tous les domaines	156 (3,3)	63 (9,86)	93 (2,28)
Méta-analyses			
Toutes	306	37	269
<i>Couvertes</i>	5 (1,63)	2 (5,41)	3 (1,11)

Au total, la presse s'est fait l'écho de 161 études scientifiques dont 65 (40,4%) impliquaient des études de la catégorie « S ». Ces études sont couvertes par un plus grand nombre d'articles de presse (11,2 articles de presse en moyenne) que les « non S » (8,2

articles de presse en moyenne). Cependant cette différence n'est pas significative. D'autre part, il ne semble pas que ce soit le facteur d'impact du journal dans lequel est publiée l'étude scientifique « S » qui influence leur sélection. En effet, celles-ci sont publiées dans des journaux scientifiques avec un facteur d'impact médian inférieur à celui des études « non S » (10,6 pour les études « S » et 17,7 pour les études « non S » différence significative, test de Student, $p=0,0003$).

1.2.3- Couverture médiatique des études initiales et des études ultérieures.

Les études initiales sont plus souvent couvertes que les études ultérieures. En effet, 13,1% des études couvertes par la presse sont des études initiales alors que seulement 2,4% des études ultérieures sont couvertes. Le facteur d'impact du journal dans lequel les études sont publiées ainsi que le type d'association étudié influencent fortement la couverture médiatique (Figure 1). En effet, si l'on considère les études « non S » (4084 études primaires), 48 des 366 études initiales sont couvertes (13,1%) contre seulement 45 des 3718 études ultérieures (1,2%). Cette différence est statistiquement significative ($\chi^2=211,52$, $p<0,0001$). Par contre, cette différence n'est plus observée pour les études « S ». En effet sur les 639 études scientifiques appartenant à cette catégorie, 39 sont des études initiales et cinq d'entre elles (12,8%) sont couvertes. En ce qui concerne les études ultérieures de cette catégorie, le même pourcentage est observé : 58 des 600 études ultérieures ont attiré l'attention des médias soit 9,7%. Compte-tenu du faible nombre (5) d'études initiales couvertes par la presse dans le cas des études « S », et du fait que les études initiales semblent être autant couvertes que les études ultérieures, nous n'avons pas fait de distinction pour les études « S ».

Sur la figure 1, on constate que dans le cas des facteurs d'impact ≤ 10 ou compris entre 10 et 30, les études initiales sont préférentiellement couvertes dans le cas des études « non S » ($\chi^2=29,11$, $p<0,0001$). Par contre, ce n'est plus le cas pour les forts facteurs d'impact (≥ 30) : les études ultérieures de la catégorie « non S » sont elles-aussi fortement couvertes ($\chi^2= 3,49$, $p=0,062$). Finalement, et quelle que soit la catégorie, le facteur d'impact du journal dans lequel l'étude est publiée influence sa couverture médiatique. Autrement dit, plus l'étude est publiée dans un journal à fort facteur d'impact et plus elle a de chance d'être couverte par la presse. En effet, lorsque l'étude est publiée dans revue prestigieuse ($IF \geq 30$), entre 46% et 72% des études scientifiques sont couvertes par la presse. A titre indicatif, 12,2% des études initiales de la catégorie « non S » sont publiées dans des journaux à facteur d'impact ≥ 30 alors que moins de 1% des études ultérieures de cette même catégorie le sont. Pour les études de type « S », 2,6% sont publiées dans des journaux à fort facteur d'impact.

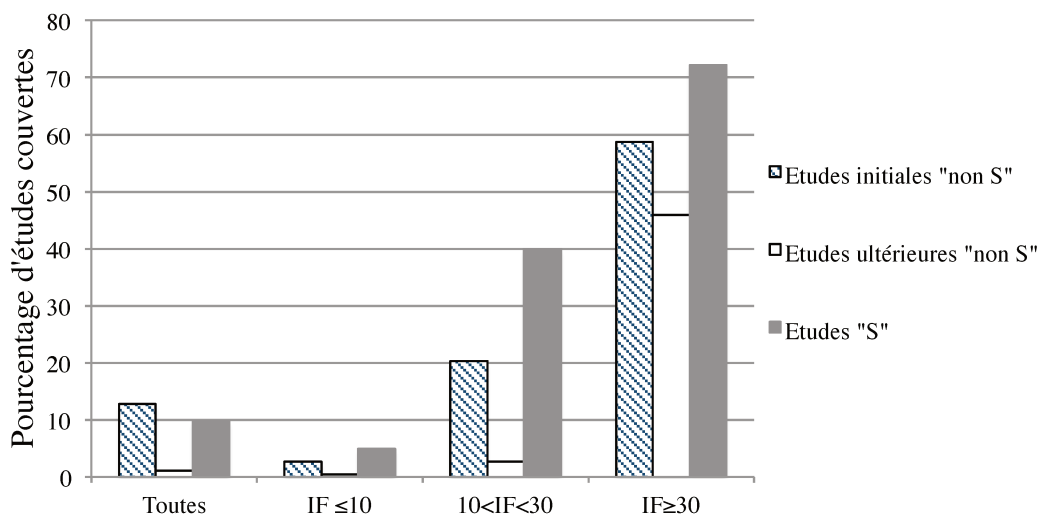


Figure 1 : Couverture médiatique des études scientifiques en fonction du type d'étude (initiales/ultérieures), de la catégorie d'étude (« S »/ « Non S ») et du facteur d'impact du journal dans lequel l'étude scientifique a été publiée.

1.2.4- Couverture préférentielle des résultats positifs

La presse préfère couvrir des résultats scientifiques initiaux et positifs (i.e. rejetant l'hypothèse nulle). En effet, sur les 156 études couvertes par la presse, 53 sont des études initiales qui rapportent toutes des effets statistiquement significatifs (Tableau 2). Cependant, 174 (43%) des études initiales contenues dans la base de données rapportaient des effets négatifs (statistiquement non significatifs). Si l'on regarde le type de résultats rapportés par les 103 études ultérieures couvertes par la presse, seulement 14 (13,6%) présentent des résultats négatifs et majoritairement dans le cadre des études de catégorie « S ». En effet, sur les 45 études ultérieures « non S » couvertes par la presse, seules cinq rapportaient un résultat négatif et n'ont été relayées que dans dix articles de presse (pour 327 articles de presse couvrant les études ultérieures « non S »). Pour la catégorie « S », neuf études rapportant des résultats négatifs ont été couvertes par 65 articles de presse. Par comparaison, les 49 autres études ultérieures rapportant un effet significatif de la catégorie « S » ont été couvertes par 668 articles de presse. Ces résultats sont présentés dans le tableau 2.

Tableau 2 : Couverture médiatique des résultats non significatifs (négatifs)

	Toutes	« S »	« Non S »
Toutes les études initiales	405	41	364
Résultats positifs	231	21	210
Résultats négatifs	174	20	154
% de résultats négatifs	43%	48,8%	42,3%
Etudes initiales couvertes			
Toutes	53	5	48
Résultats négatifs	0	0	0
Etudes ultérieures couvertes			
Toutes	103	58	45
Résultats négatifs	14	9	5
Articles de presse			
Etudes ultérieures	1022	695	327
Couvrant des résultats négatifs	75	65	10

Si l'on considère que la proportion de résultats négatifs compris dans la base de données est la même que celle observée pour les études initiales (43%), ces résultats montrent que, non seulement la presse s'en fait peu l'écho, mais aussi que lorsqu'elle le fait, elle y consacre peu d'articles.

On observe le même type de résultats pour les méta-analyses. Sur les cinq méta-analyses couvertes par la presse, une seule présente des résultats négatifs qui ne sont relayés que par quatre articles de presse. Les quatre autres méta-analyses rapportant des résultats positifs ont elles été couvertes par 82 articles de presse.

1.2.5- Reproductibilité des études scientifiques couvertes par la presse.

Nous avons ensuite analysé la validité des études scientifiques médiatisées. Pour cela, nous avons déterminé les taux de reproductibilité des études scientifiques couvertes par la presse en fonction des domaines de recherche, du type d'étude (initiale/ultérieure) de la catégorie d'étude (« S »/ « Non S ») et du facteur d'impact du journal dans lequel l'étude a été publiée. Les résultats sont représentés sur les figures 2 (en fonction du type et des catégories d'études) et 3 (en fonction du domaine de recherche).

Le taux de reproductibilité des 156 études couvertes par la presse est de 48,7% (76 études ont été considérées comme reproduites lorsque leurs résultats ont été comparés à ceux des méta-analyses). Si l'on analyse la reproductibilité en fonction du type d'étude, on constate que la reproductibilité des études initiales est inférieure à celle des études ultérieures ($\chi^2= 6,99$, $p=0,0082$). En effet, 53 des études « S » et « non-S » couvertes par la presse sont des études initiales et leur taux de reproductibilité est de 34% alors que celui des études ultérieures est de 56,3%. Les études de type « S » ont un taux de reproductibilité de 49,2% et celui-ci n'est pas influencé par le facteur d'impact dans lequel l'étude a été publiée (Figure 2). Seuls 33,3% des études initiales « non-S » sont reproduits et ce taux n'est pas non plus influencé par le facteur d'impact du journal scientifique (Figure 2). Par contre, le taux de reproductibilité des études ultérieures « non S » est supérieur (64,4%). La différence est significative si l'on compare leur reproductibilité à celle des études initiales « non S » (33,3% ; $\chi^2= 9,0$, $p=0,0027$). D'autre part, leur taux de reproductibilité augmente avec le facteur d'impact du journal dans lequel elles ont été publiées ($p=0,0055$). Finalement, les résultats présentés dans journaux prestigieux ne sont pas plus valides si l'on considère les études initiales de type « non S » et les études de « S ». Par contre, ceux des études ultérieures de type « non S » sont d'autant plus fiables qu'ils sont publiés dans des journaux à fort facteur d'impact (≥ 30).

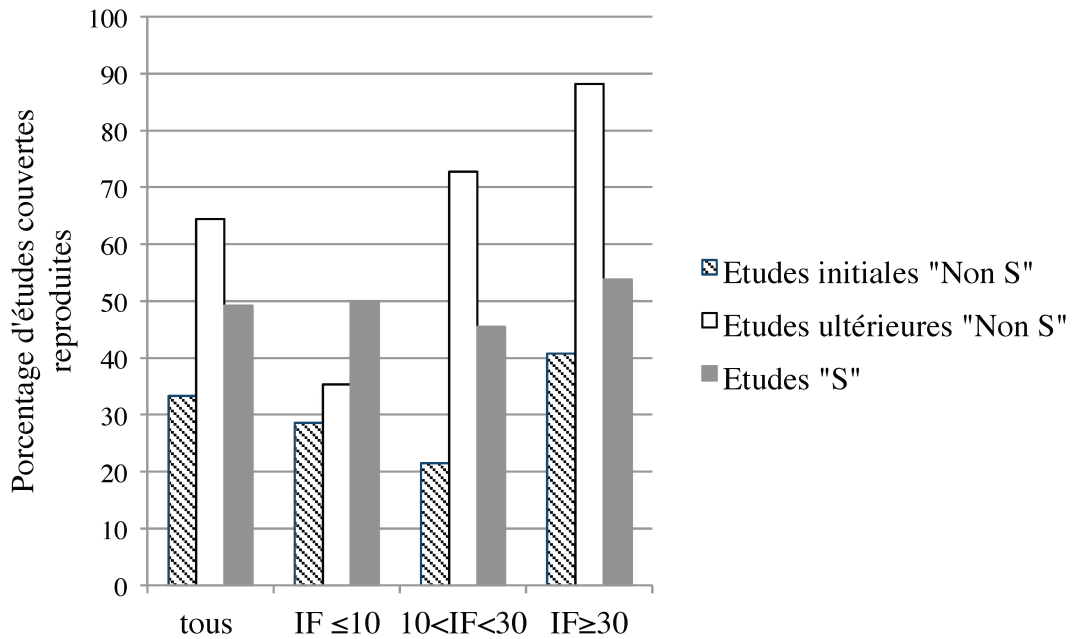


Figure 2 : Taux de reproductibilité des études scientifiques couvertes par la presse en fonction du type d'étude (initiale/ultérieure), de la catégorie d'étude (« S »/ « non S ») et du facteur d'impact du journal.

Enfin, nous avons analysé le taux de reproductibilité des études couvertes par la presse en fonction des domaines de recherche (Figure 3).

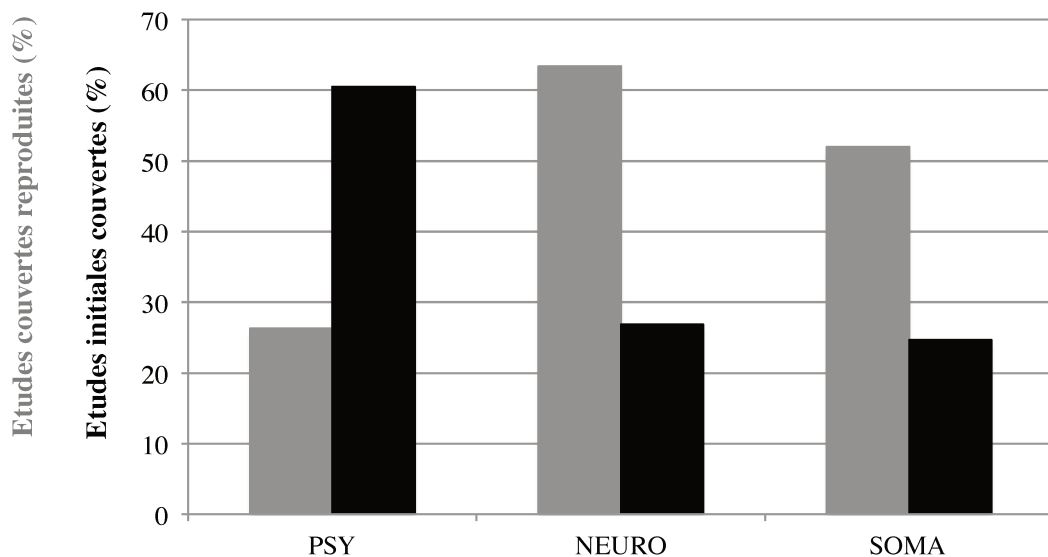


Figure 3 : Taux de reproductibilité des études scientifiques couvertes par la presse en fonction du domaine biomédical. PSY : psychiatrie, NEURO : neurologie et SOMA : maladies somatiques. On retrouve aussi le pourcentage d'études initiales dans les études couvertes par la presse pour chaque domaine.

Les taux de reproductibilité dans le domaine de la psychiatrie sont inférieurs à ceux des études en neurologie (26,3% contre 63,41% ; $\chi^2= 10,94$, $p=0,0009$) et à ceux des études somatiques (26,3% contre 51,9% ; $\chi^2= 6,8$, $p=0,0091$). Ce faible taux de reproductibilité

semble être corrélé à la couverture préférentielle des études initiales dans le domaine de la psychiatrie (Figure 3). En effet, 60,5% des études scientifiques couvertes par la presse dans le domaine de la psychiatrie sont des études initiales contre 26,8% pour la neurologie ($\chi^2= 9,13$, $p=0,0025$) et 24,7% pour les maladies somatiques ($\chi^2= 14,11$, $p=0,0002$). Nous avons montré que les études initiales étaient peu reproduites (Dumas-Mallet et al. 2016). Ceci explique en partie que le taux de reproductibilité des études couvertes par la presse dans le domaine de la psychiatrie soit inférieur à ceux observés pour la neurologie et les maladies somatiques.

1.2.6- La presse couvre-t-elle les résultats d'études contredisant ceux d'études qu'elle a précédemment couvertes ?

Le taux de reproductibilité des études couvertes par la presse est globalement faible (48,7%). Ceci implique que plus de la moitié des résultats de la recherche biomédicale présentés dans la presse sont soit exagérés soit faux. Nous avons donc regardé si la presse mentionnait ces résultats contradictoires. Nous avons analysé, pour chaque étude couverte et par la suite réfutée ou atténuée, si la presse relayait ces informations. Nous n'avons relevé qu'un seul cas pour lequel la presse a mentionné la réfutation d'une étude qu'elle avait précédemment couverte. L'étude initiale publiée dans *Science* (IF=31) avait été couverte par 50 articles de presse différents. Par comparaison, la méta-analyse, qui réfute l'étude initiale a été publiée dans *JAMA* (IF=30) mais n'a été couverte que par quatre articles de presse.

Une hypothèse souvent avancée pour expliquer ce manque de suivi de la presse est que les études ultérieures sont publiées dans des journaux moins prestigieux et sont donc moins visibles (Holtzman et al. 2005). Nous avons analysé les 104 associations²⁰ distinctes présentées dans les médias. Nous avons séparé les associations qui étaient confirmées (47 associations soit 45,2%). Il reste donc 57 associations qui ont été atténuées ou réfutées par les méta-analyses. Pour chacune de ces associations, nous avons compilé les résultats des études primaires qui étaient publiées dans des journaux avec un facteur d'impact >10. Nous avons alors distingué trois cas : (1) au moins une étude publiée dans un journal avec un IF>10 contredisait l'étude couverte par la presse et pouvait donc être considérée comme visible, (2) aucune étude n'était publiée dans des journaux avec des IF>10 ou (3) aucune étude publiée dans un journal avec un IF>10 ne contredisait l'étude précédemment couverte par la presse. Les résultats sont présentés dans le tableau 3.

Pour 27 des associations couvertes par la presse, aucune publication invalidant l'étude précédemment couverte n'était visible pour les journalistes. Par contre, pour 30 des

²⁰ Les associations correspondent aux sujets sur lesquels portent les méta-analyses. Par exemple l'association entre l'exposition aux pesticides et la maladie de Parkinson. Pour certaines de ces associations (34), plusieurs études primaires sont couvertes. Ceci explique qu'il y ait 156 études couvertes mais seulement 104 associations.

associations (52,6%), une étude ultérieure contredisant l'étude couverte par la presse était publiée dans un journal avec un facteur d'impact >10. De plus, 22 d'entre elles étaient publiées dans des journaux avec des facteurs d'impact supérieurs ou au moins égaux à ceux de l'étude précédemment couverte. Cependant, la presse n'a fait mention que d'une seule réfutation et n'y a consacré que quatre articles. Autrement dit, pour 21 des 57 associations, il existait au moins une étude ultérieure contredisant l'étude déjà couverte et avec une visibilité au moins égale à celle-ci. On voit donc que le manque de visibilité des études ultérieures n'explique que partiellement le fait que la presse ne couvre que rarement les études contredisant une étude préalablement couverte.

Tableau 3 : Répartition des 104 associations couvertes par la presse en fonction de leur reproductibilité et de la visibilité des études ultérieures invalidant le résultat couvert.

	104 associations			
	Confirmées	Atténuées ou Réfutées		
	47	57		
	45,2%	(1)	(2)	(3)
IF>10		30 (52,6%)	17	10
IF> étude couverte		22	NA	NA
Couverture réfutation		1	0	0

1.3- Apports et limites de cette étude

1.3.1- les apports

Nos résultats montrent que la presse favorise la couverture des études initiales. En effet, 13,1% des études initiales contenues dans la base de données sont couvertes (53/405) alors que seulement 2,4% des études ultérieures le sont (103/4218). Elle couvre aussi préférentiellement les études publiées dans des journaux à fort facteur d'impact (entre 45 et 72% des études publiées dans des journaux avec $IF \geq 30$ sont couvertes). Nous avons également confirmé que la presse privilégiait les études impliquant le mode de vie (Bartlett et al. 2002; Stryker 2002). Ces études présentent les résultats d'association entre des pathologies et des facteurs de risques sur lesquels le patient peut agir. Par comparaison, les études « non S », qui compilent des résultats d'imagerie cérébrale, de facteurs génétiques ou d'autres facteurs contre lesquels le patient ne peut se prémunir, sont moins souvent couverts. Cette sélection préférentielle des études « S » confirme l'objectif principal des journalistes : être lus (Laurance 1998). Enfin, notre analyse confirme que la presse couvre presque exclusivement des résultats positifs. En effet, aucun des résultats d'études initiales statistiquement négatifs n'a été couvert et très peu pour les études ultérieures. Seulement 5% des articles de presse y sont dédiés.

L'objectif principal de cette étude était de mesurer la validité des études couvertes par la presse. Nous avons donc mesuré le taux de reproductibilité : plus de la moitié des études scientifiques rapportées dans la presse (51,3%) ne sont pas confirmées par les études

ultérieures sur le même sujet. Ce taux de reproductibilité varie en fonction de plusieurs facteurs. Tout d'abord, en fonction du type d'étude : les études initiales couvertes sont moins souvent reproduites (34%) que les études ultérieures (56,3%). Ce résultat est en accord avec les observations qui sont faites dans la littérature scientifique sur la faible reproductibilité des études initiales comparées aux études ultérieures sur un même sujet (Dumas-Mallet et al. 2016; Ioannidis et al. 2001b; Ioannidis et al. 2011c). Ceci explique le manque de reproductibilité des études scientifiques couvertes par la presse : en effet celle-ci privilégie les études initiales. Pour les domaines de recherche, les études en psychiatrie sont moins souvent reproduites que celles en neurologie ou pour les maladies somatiques. Ceci est en partie lié au fait que 60% des études médiatisées en psychiatrie sont des études initiales. Comme celles-ci souffrent d'une faible reproductibilité, il n'est pas étonnant que leur reproductibilité dans les médias soit moindre. Enfin, nous n'avons pas observé d'effet du facteur d'impact du journal dans lequel l'étude est publiée sauf dans le cas des études ultérieures de type « non S ». Autrement dit, mis à part ce cas particulier, les études publiées dans des journaux avec des $IF \geq 30$ et couvertes par la presse ne sont pas plus fiables que les celles publiées dans des journaux moins prestigieux.

Nous avons également analysé la couverture éventuelle d'études contredisant des études précédemment couvertes. En effet, compte-tenu du faible taux de reproductibilité, il semblerait important que la presse informe le public de ces invalidations. Nous n'avons observé qu'un seul cas où la presse s'est fait l'écho d'une réfutation en mentionnant explicitement l'étude initiale précédemment couverte. Il est néanmoins vraisemblable que la presse se soit fait l'écho de cette invalidation parce que la méta-analyse était elle aussi publiée dans un journal scientifique prestigieux et donc visible. Cependant, nous avons observé que dans 52,6% des cas, une étude ultérieure invalidant l'étude précédemment couverte avait été publiée dans un journal avec un facteur d'impact supérieur à 10. L'hypothèse selon laquelle la presse ne peut pas faire état des réfutations parce qu'elles ne sont pas visibles n'est pas valable ici (Holtzman et al. 2005).

1.3.2- Les limites

Nous avons limité notre étude de la couverture médiatique des résultats de la recherche à une seule association par article scientifique. En effet, dans le cas des études génétiques, plusieurs polymorphismes étaient étudiés dans une même publication. Dans cette situation, nous n'avons retenu que la plus crédible de ces associations, c'est à dire celle qui était statistiquement la plus significative. Il est donc probable que nous ayons surestimé le taux de reproductibilité des études scientifiques couvertes par la presse.

Notre approche utilise les résultats de méta-analyses pour nous permettre de déterminer la validité des études primaires couvertes par la presse. Cependant, de nombreuses études d'association n'ont pas fait l'objet de méta-analyses et nous n'avons donc pas analysé leur couverture médiatique. Nous ne pouvons donc pas généraliser ces résultats à ce type d'étude.

Nous avons également limité nos recherches aux articles de presse écrits dans le mois suivant la publication de l'article scientifique. D'autres articles de presse, publiés ultérieurement ont probablement mentionné les résultats de l'étude considérée, mais il devenait alors difficile de s'assurer du lien avec l'article scientifique en question.

Nous avons montré que la presse favorisait la couverture des études initiales et des études impliquant le mode de vie. Nous avons montré que le taux de reproductibilité des études initiales de type « non S » était inférieur à celui des études ultérieures. Cependant nous n'avons pas pu apporter de conclusion concernant spécifiquement la reproductibilité des études initiales de type « S » puisque nous n'avons identifié que cinq études de ce type couvertes par la presse.

Enfin, nous avons choisi d'étudier la couverture médiatique des résultats scientifiques dans la presse alors que, pour 67% des européens, la télévision est la principale source d'information (Eurobaromètre 2007). Cependant, nous voulions quantifier la reproductibilité des études scientifiques couvertes par les médias. Ceci n'est possible que si l'étude scientifique couverte est identifiable. Or, les programmes de télévision présentent très rarement les résultats d'études scientifiques précises (Bourdaa et al. 2015).

1.4- Conclusion

Notre étude montre que la presse privilégie la couverture d'études scientifiques initiales présentant des résultats positifs. Elle ne s'intéresse quasiment jamais aux résultats statistiquement non significatifs. Comme elle favorise la couverture de résultats initiaux, ces résultats sont invalidés dans plus de la moitié des cas. De plus, la presse n'informe que rarement le public de ces invalidations. Ceci avait déjà été évoqué dans un commentaire de la revue *Nature Biotechnology* (Bubela et al. 2009). Les résultats présentés ici permettent d'étayer ce commentaire par des observations quantitatives.

Cette préférence pour les résultats initiaux peut être expliquée par le fait que ces résultats sont souvent publiés dans des journaux à fort facteur d'impact. Les journalistes utilisent fréquemment ces journaux comme sources (Nelkin 1996), leur accordent un crédit important (Hansen 1994) et les croient désintéressés (Wilkie 1996). Cependant des auteurs ont suggéré que ces journaux prestigieux favorisaient les résultats spectaculaires et susceptibles d'attirer l'attention des médias (Bucchi 2015; Chew et al. 2007). Les éditeurs de

ces journaux sont donc loin d'être désintéressés. Ces résultats montrent que les journalistes devraient être plus prudents et s'assurer de la validité des résultats présentés dans ces journaux scientifiques prestigieux. Lorsqu'ils couvrent une étude initiale, ils devraient s'efforcer d'informer le public que ces résultats sont incertains et qu'ils doivent être confirmés par des études ultérieures.

II- Présentation des résultats de la recherche biomédicale dans la presse grand public

La connaissance scientifique résulte d'un processus cumulatif au cours duquel des résultats initiaux prometteurs mais incertains sont reproduits par d'autres scientifiques pour arriver à un consensus. Un résultat initial est donc par définition incertain et il est parfaitement normal que tous les résultats initiaux ne puissent être répliqués (Open Science Collaboration 2015). De fait, plusieurs travaux, dont le notre (chapitre II), ont montré que la majorité des résultats d'études initiales étaient invalidés ou fortement atténués par des études ultérieures. Un article d'opinion affirme, sans apporter de données quantitatives, que les médias couvrent préférentiellement des études initiales rapportant des résultats positifs et exagèrent souvent les perspectives d'application (Bubela et al. 2009). Un autre article d'opinion, lui, souligne que les études initiales dans le domaine biomédical sont incertaines et ces auteurs vont même jusqu'à recommander que les journalistes s'abstiennent de les couvrir (Schwartz et al. 2004). Comme nous venons de le voir, notre travail apporte des données quantitatives robustes à l'appui de ces opinions : les journaux privilégient les études initiales rapportant des effets positifs et publiés dans des revues prestigieuses alors même que leurs résultats sont souvent invalidés par les méta-analyses.

Dans notre revue de la littérature (chapitre I), nous avons présenté de nombreuses études en science de la communication qui traitent de l'incertitude scientifique. Cependant, l'incertitude étudiée dans ces cas se résume à celle liée aux risques potentiels associés à la science décrite (National Academies of Sciences 2017). L'incertitude inhérente aux études initiales est ignorée. Cependant, elle pose problème (Bubela et al. 2009; Schwartz et al. 2004) et représente une autre forme d'incertitude. Sa prise en compte par les médias est une question importante qui n'a pratiquement pas été abordée par les chercheurs en sciences de la communication, si ce n'est l'étude proposée par Holtzman (Holtzman et al. 2005).

Dans cette étude les auteurs analysent la couverture médiatique de 24 découvertes génétiques associées à des pathologies humaines (Holtzman et al. 2005). Dans une autre étude, les mêmes auteurs avaient interviewé 16 journalistes scientifiques et 79% d'entre eux ont affirmé que la couverture d'une étude devait impérativement préciser son statut : a-t-elle

ou non été répliquée (Mountcastle-Shah et al. 2003)? En l'absence de répllication connue, il était essentiel, pour ces journalistes, de préciser que l'étude en question était préliminaire et devait être validée par des études ultérieures. Malheureusement, ces bonnes intentions étaient assez rarement suivies. Le statut de l'étude ou la nécessité de la valider par des réplifications n'étaient mentionnés que dans un tiers des 228 documents médiatiques examinés (Holtzman et al. 2005). Cependant, dans leur analyse de la couverture médiatique de 24 études génétiques, ces auteurs ne précisent pas lesquelles étaient effectivement des études initiales et lesquelles ont été invalidées par les études ultérieures.

Nous avons donc analysé la façon dont les journalistes présentaient l'incertitude et les résultats de la recherche dans 426 articles de presse couvrant les résultats de 40 études initiales. Les études scientifiques sont issues de la base de données que nous avons créée (chapitre II) et les articles de presse ont été identifiés selon le protocole décrit dans le paragraphe I. Nous connaissons la validité de chacune des études couvertes par la presse. Nous avons analysé plusieurs facteurs. Tout d'abord nous avons classé les articles de presse en fonction de leur titre : est-il exagéré ou non ? Ensuite nous avons étudié le contenu de chaque article de presse et recherché la présence de trois éléments. Premièrement, l'étude scientifique est-elle présentée comme une étude initiale ? Deuxièmement, l'article mentionne-t-il l'importance de reproduire les résultats ? Troisièmement, l'article contient-il une affirmation exagérant la robustesse des résultats? Nous avons identifié les auteurs de ces mentions : les scientifiques qui ont réalisé ces études, d'autres experts du domaine ou encore le journaliste. Ces résultats sont décrits dans un projet de publication actuellement soumis à la revue *Science communication* (Dumas-Mallet et al. 2017soumis). Cet article est en annexe 3.

2.1- Méthodologie

2.1.1- Sélection des articles de presse

Nous avons analysé le contenu de 426 articles de presse différents qui couvraient les résultats de 40 études initiales. Les études scientifiques ont été identifiées dans la base de données que nous avons décrite dans le chapitre II. Nous avons compilé les résultats de méta-analyses répertoriées sur *PubMed* et publiées entre 2008 et 2012. Ces méta-analyses concernaient l'association entre des facteurs de risque et des pathologies appartenant à trois domaines de la recherche biomédicale : la psychiatrie, la neurologie et quatre autres maladies somatiques (Dumas-Mallet et al. 2016). Nous avons utilisé la base de données *Dow Jones Factiva* pour identifier les études scientifiques (initiales, ultérieures et méta-analyses) couvertes par la presse. Nous avons ainsi identifié 156 études rapportées dans 1475 articles de presse. Ces résultats ont été présentés dans la première partie de ce chapitre. Nous avons ici

analysé le contenu des 426 articles de presse se rapportant aux études initiales qui avaient fait l'objet d'au moins deux articles de presse (40 études sur les 53 études initiales couvertes). Certaines publications scientifiques ultérieures à ces 40 études initiales et portant sur le même sujet ont aussi attiré l'attention de la presse. Par exemple dans le cas de la corrélation entre l'âge du père et le risque d'autisme chez l'enfant. Dans ce cas, comme pour 9 autres des études initiales, la couverture médiatique des études ultérieures, soit 111 articles de presse, a aussi été analysée à titre de contrôle.

2.1.2- Analyse du contenu des articles de presse

Pour cette étude nous avons analysé 3 éléments : le « ton » du titre de l'article, la présentation de l'étude comme initiale et la description de l'incertitude ou de la robustesse des résultats.

Les titres de chaque article de presse ont été classés par deux auteurs (EDM et FG). Deux catégories ont été définies : des titres exagérés ou des titres neutres. Les titres ont été classés comme exagérés si (1) des mots tels que « *breakthrough* »²¹ ou « *key finding* » étaient utilisés ou si (2) les implications thérapeutiques étaient proposées alors que les chercheurs avaient simplement identifié un marqueur génétique d'une pathologie. Par contre, lorsque les titres mentionnaient des mots tels que « *study* », « *researchers* » ou « *research* », ils ont été classés comme neutres. Enfin, les titres factuels comme « *genetic variations found in depression* » ont été aussi classés comme neutres. Quelques exemples sont présentés dans le tableau 4.

Tableau 4 : Exemples de titres exagérés et de titres neutres.

Titres exagérés	Titres neutres
<i>Genetics may hold key to curing misunderstood neural disorder</i>	<i>Genetic defect studied</i>
<i>Breakthrough claimed for attention disorder</i>	<i>Scan may spot attention deficit disorder</i>
<i>Older mothers "lead to autism"</i>	<i>Autism in children is linked to older mothers</i>
<i>The depression gene</i>	<i>Study links genes to stress, depression</i>
<i>Insanity linked to size of the brain</i>	<i>Schizophrenia linked to brain abnormalities</i>
<i>Secret code behind MS finally cracked</i>	<i>Researchers find genes linked to MS</i>
<i>Anti-inflammatory drugs ward off Parkinson's</i>	<i>Painkillers may reduce the risk of Parkinson's</i>
<i>Magic key to breast cancer fight</i>	<i>Scientist pinpoint genes that raise your breast cancer risk</i>
<i>Two more markers in road to cure</i>	<i>New cancer clues</i>

Pour s'assurer de la validité de cette classification, le codage a été réalisé indépendamment par un troisième auteur (AS). Les résultats des deux codages ont été comparés en utilisant le coefficient Kappa de Cohen (Sim et al. 2005). Les désaccords entre codeurs ont été résolus après discussion.

²¹ Les mots sont indiqués en anglais puisqu'il s'agit d'une analyse de contenu réalisée sur des articles rédigés en anglais. Nous avons jugé que la traduction n'avait ici pas de sens.

Pour chaque article de presse, nous avons recherché des expressions qui montraient que les journalistes avaient correctement perçu et donc présenté les études comme des études initiales. En effet, les journalistes ont tendance à présenter tous les résultats de la recherche comme des résultats révolutionnaires (Nelkin 1995). Nous avons donc tenté de différencier les mots qui identifiaient comme telles de « vraies » études initiales.

Par exemple, nous avons considéré que « *for the first time* » et « *in a preliminary survey* » décrivaient des études initiales alors que des qualificatifs tels que « *the new finding* » étaient trop imprécis (Tableau 5). Afin de vérifier la fiabilité de notre classification, nous avons réalisé la même analyse sur les 111 articles de presse couvrant les 10 études ultérieures sur le même sujet. Nous avons ainsi contrôlé l'utilisation des qualificatifs que nous considérons comme déterminants pour juger que le journaliste avait effectivement présenté l'étude comme initiale.

Tableau 5 : Exemples de qualificatifs reflétant que le journaliste perçoit l'étude comme une étude initiale.

Qualificatifs déterminants	Qualificatifs imprécis
<i>The first study to show a measurable...</i>	<i>The findings were based on a survey...</i>
<i>For the first time...</i>	<i>A new study that indicates...</i>
<i>Initial test...</i>	<i>A major advance...</i>
<i>The finding is still preliminary...</i>	<i>The breakthrough will have major...</i>
<i>Scientists have discovered a gene...</i>	<i>A breakthrough in brain scan technology...</i>
<i>Researchers have uncovered another gene...</i>	<i>Researchers found that...</i>
<i>The discovery, which links the disorder to a gene that regulates...</i>	<i>A discovery that may help scientists to develop new treatments.</i>
<i>The discovery of this gene variant...</i>	<i>A discovery that could lead to earlier diagnosis</i>
<i>The first convincing evidence...*</i>	<i>Researchers have identified a gene linked...</i>

*Dans les cas où cette expression était associée dans le même article de presse avec la mention que l'étude en question reproduisait les résultats d'études précédemment réalisées, nous avons considéré que le mot « *first* » qualifiait alors le mot « *convincing* » et non le mot « *evidence* ». Dans ces cas, nous avons conclu que l'étude n'était pas décrite comme une étude initiale.

Nous avons aussi analysé le contenu des articles de presse pour déterminer si les journalistes tenaient compte de l'incertitude inhérente aux résultats initiaux ou si au contraire ils les présentaient comme plus robustes qu'ils ne l'étaient. Ainsi, la mention de l'importance de reproduire les résultats est une indication de l'incertitude qui leur est liée. A l'inverse, nous avons recherché des qualificatifs suggérant la fiabilité des résultats. Plusieurs exemples sont présentés dans le tableau 6.

Tableau 6 : Exemples d'expressions qui soulignent l'incertitude ou exagèrent la fiabilité.

Incertitude- Nécessité de reproduire	Fiabilité des résultats
<i>Replication of the results is absolutely essential</i>	<i>The initial data is robust</i>
<i>Their results must be replicated</i>	<i>It provides rigorous confirmation of a link</i>
<i>Tests on larger populations of adults must be performed</i>	<i>It establishes which gene plays an important role</i>
<i>[The data] need to be verified by further studies</i>	<i>A proven direct genetic link</i>
<i>If the work stands up</i>	<i>A clear-cut chemical abnormality</i>
<i>More work is needed to confirm the findings.</i>	<i>These results demonstrate</i>
<i>The study needs to be repeated,</i>	<i>This kind of evidence makes it clear</i>
<i>They cautioned that the number of patients tested was small</i>	<i>This is the first demonstration of a gene affecting a brain response</i>
<i>It will be important for other scientists to confirm that.</i>	<i>The statistical significance is the highest</i>
<i>The study is preliminary,</i>	<i>Irrefutable evidence</i>

2.2- Résultats

2.2.1- Validité des codages

La classification des titres de presse en titres neutres ou exagérés est en partie subjective. C'est pourquoi le codage a été réalisé tout d'abord par 2 auteurs (EDM et FG) puis comparé au codage réalisé par un 3^{ème} auteur (AS). Le coefficient Kappa de Cohen a été calculé. Il est élevé (0,71) ce qui dénote un bon accord entre les codeurs et donc la robustesse de la classification.

Nous avons choisi des qualificatifs qui nous permettaient de penser que les journalistes avaient perçu et décrit les études comme initiales. Pour vérifier ces choix nous avons comparé les résultats de deux catégories d'articles de presse : ceux couvrant des études initiales (426 articles de presse) et ceux couvrant des études ultérieures sur le même sujet (111 articles de presse). En utilisant les qualificatifs que nous avons définis, nous avons compté 243 articles de presse sur les 426 qui présentaient avec raison les études comme des études initiales (57%). Par contre, nous n'avons identifié que 14 articles de presse sur les 111 qui présentaient à tort l'étude comme une étude initiale (11%). Cette différence bien que statistiquement significative (test du Chi² : $\chi^2=69,7$, $p<0,0001$) montre que nos critères de définition ne sont pas parfaitement discriminatifs. Nous ne pouvons pas affirmer que les journalistes, lorsque qu'ils emploient des qualificatifs tels que « *for the first time* » ou « *the discovery of a gene* », les utilisent à bon escient. En effet, leur volonté de présenter les résultats comme des découvertes et de décrire la recherche comme une « course » entre les scientifiques (Friedman et al. 1999), peut les conduire à présenter n'importe quel résultat comme initial pour le rendre sensationnel.

Enfin, les derniers codages réalisés concernent la présence ou non d'expressions indiquant la robustesse ou au contraire l'incertitude liée aux résultats scientifiques présentés par les journalistes ainsi que ceux qui les ont formulés (scientifiques auteurs de l'étude,

experts du domaine ou journalistes). Le nombre d'articles qui incluait des citations des auteurs de la publication scientifique a également été quantifié. Ces codages ont été réalisés indépendamment par deux des auteurs (EDM et FG). Les résultats de ces codages étaient très similaires et les rares différences ont été résolues par discussion entre les deux auteurs.

2.2.2- Description des articles de presse couvrant les études initiales

Le nombre moyen d'article de presse couvrant chacune des études initiales que nous avons analysé ici (40 études) est de 11. Le nombre d'articles de presse varie de manière importante puisque certaines études ne sont couvertes que par deux articles alors que d'autres sont couvertes par 50 articles de presse indépendants. Le contenu des articles couvrant une même étude était souvent assez similaire et reprenait parfois en grande partie les textes des dépêches publiées par les agences de presse *Reuters* ou *Associated Press*.

La longueur moyenne des articles de presse est de 417 mots avec là encore des variations fortes : certains articles ne contiennent que 17 mots alors que d'autres atteignent 1750 mots.

Le taux de reproductibilité des résultats de cette sous-population d'études initiales est de 32,5% (13 des études sont confirmées par les méta-analyses). Il est similaire à celui observé pour l'ensemble des études initiales de la base de données (chapitre II). Près des trois quarts des articles de presse sélectionnés ici couvrent les résultats d'études génétiques (23 études et 304 articles de presse). Les études dans les domaines de la psychiatrie (16 études et 166 articles) et des maladies somatiques (16 études et 161 articles) sont plus facilement couvertes que celles en neurologie (8 études et 99 articles).

Nous avons constaté que les articles de presse couvrant une même étude présentaient des titres toujours différents (voir les documents en annexe 3). Nous avons classifié 150 titres comme exagérés (35,2%). Nous avons quantifié 243 articles de presse dans lesquels l'étude était présentée avec raison comme une étude initiale (57%). Les auteurs des publications scientifiques étaient cités dans 70,9% des cas (302 articles de presse). Enfin, nous avons identifié 91 articles dans lesquels la nécessité de reproduire les expériences était indiquée (21,4%) et 105 pour lesquels la fiabilité des résultats était exagérée (24,6%). Dans la majorité des cas, ce sont des scientifiques (auteurs de l'étude couverte ou des experts du domaine) qui ont souligné l'importance de reproduire les résultats pour pouvoir les confirmer (81,3% des cas) ou exagéré la robustesse des résultats (78%).

2.2.3- Analyse des corrélations entre la description d'une étude comme initiale, la mention de l'incertitude ou de la fiabilité des résultats.

Nous avons tout d'abord testé si, lorsque l'article de presse décrivait explicitement l'étude comme une étude initiale, cette description était associée à la nécessité de reproduire les résultats de l'étude pour la valider (Tableau 7). Nous avons observé que la mention de la nécessité de reproduire les résultats était plus souvent présente lorsque l'étude était décrite comme une étude initiale (26,3%) que dans le cas contraire (14,7%). Cette différence est statistiquement significative (test du Chi² : $\chi^2=8,3$, $p=0,0039$). Nous avons également testé si l'affirmation de la robustesse des résultats était contrebalancée par la mention de la nécessité de reproduire les résultats. Les deux mentions sont associées dans 31,9% des cas. Dans 22,7% des cas l'affirmation de la robustesse des résultats n'est pas contrebalancée. Cette différence n'est cependant pas significative (test du Chi² : $\chi^2=3,25$, $p=0,071$).

Tableau 7 : Corrélations entre les critères analysés dans les articles de presse

Etude initiale et Nécessité de reproduire				Nécessité de reproduire et Robustesse			
0-0	0-1	1-0	1-1	0-0	0-1	1-0	1-1
156	27	179	64	259	76	62	29
	14,7%		26,3%		22,7%		31,9%
Citations des Auteurs et Nécessité de reproduire				Citations des Auteurs et Nécessité de reproduire (articles >200 mots)			
0-0	0-1	1-0	1-1	0-0	0-1	1-0	1-1
114	10	221	81	59	6	188	76
	8,1%		26,8%		9,2%		28,8%
Exagération du titre et Nécessité de reproduire				Exagération du titre et robustesse			
0-0	0-1	1-0	1-1	0-0	0-1	1-0	1-1
201	75	134	16	204	72	117	33
	27,2%		10,7%		26%		22%

0 indique que le critère recherché (étude initiale, nécessité de reproduire, robustesse des résultats ou citation des auteurs) n'est pas présent. 1 indique que le critère recherché est présent.

2.2.4- Influence des scientifiques sur la mention de la nécessité de reproduire les résultats

Dans près de la moitié des cas, ce sont les auteurs des études qui mentionnent l'importance de reproduire les résultats pour les valider (44/91). On pourrait donc penser que, lorsque les articles de presse ne mentionnent pas la nécessité de reproduire les résultats, les journalistes n'ont pas cité les auteurs de la recherche. Nous avons donc testé cette hypothèse. Les résultats montrent que lorsque les auteurs ne sont pas cités dans l'article de presse, la mention de la nécessité de reproduire est moins présente (9,2% contre 28,8%). Cette différence est significative (test du Chi² : $\chi^2=18,4$, $p<0,0001$). Cependant, ce résultat pourrait être expliqué par la longueur des articles de presse. En effet, on peut penser que si l'article est court, le journaliste aura moins tendance à inclure une citation des auteurs et donc la mention

de la nécessaire répliation. Nous avons donc réalisé le même test mais sur des articles d'au moins 200 mots (320 articles de presse). Là encore, nous avons retrouvé la même corrélation : la nécessité de reproduire les résultats est plus souvent mentionnée lorsque les auteurs sont cités dans l'article. On voit donc que les auteurs des études sont souvent responsables de la mention de la nécessité de reproduire les résultats mais qu'ils sont loin de le faire systématiquement. En effet, 264 des 426 articles de presse incluaient une citation des auteurs et parmi ces 264 la mention d'une nécessaire répliation n'était présente que dans 76 articles soit 28,8%.

2.2.5- L'exagération des titres est inversement corrélée à la mention de la nécessité de reproduire les résultats

Notre recherche confirme que la presse utilise souvent des qualificatifs hyperboliques pour décrire les résultats scientifiques, notamment dans les titres des articles (Caulfield 2004; Caulfield et al. 2012). Nos deux hypothèses sont donc que lorsque le titre d'un article de presse est exagéré, on retrouvera (1) plus souvent dans le texte de l'article une affirmation de la fiabilité des résultats et (2) moins souvent une mention de la nécessité de les reproduire. Nous n'avons pas validé la première hypothèse : il n'y a pas de corrélation entre l'exagération du titre et l'affirmation de la robustesse des résultats (Tableau 4). En revanche, nos données (Tableau 7) apporte un solide crédit à la deuxième hypothèse : les articles de presse dont les titres sont exagérés contiennent moins de mentions de la nécessité de répliquer (test du Chi² : $\chi^2=15,8$, $p<0,0001$).

2.2.6- Influence du taux de reproductibilité

Nous avons ici testé la corrélation entre la reproductibilité - *a posteriori* - d'une étude et la mention de la nécessité de reproduire les résultats ou de la robustesse de cette étude. Cette analyse peut paraître peu logique. En effet, lorsque les journalistes couvrent une étude initiale, ils ne peuvent pas connaître sa reproductibilité puisque ni les études ultérieures, ni les méta-analyses n'ont encore été réalisées. Cependant cette analyse a révélé une corrélation étonnante. Si la présence ou non d'affirmation de la robustesse des résultats semble indépendante de la reproductibilité ou non d'une étude, ce n'est pas le cas pour la mention de la nécessité de reproduire les résultats. En effet, sur les 91 mentions d'une nécessaire répliation, seulement quatre d'entre elles étaient exprimées pour des études par la suite validées. Ces mentions étaient donc très majoritairement présentes dans les articles de presse couvrant les 27 études initiales qui ont par la suite été contredites (test du Chi² : $\chi^2=60,2$, $p<0,0001$).

2.2.7- Influence de l'année de publication de l'étude scientifique

Depuis maintenant plusieurs années et comme nous l'avons déjà largement décrit, la littérature biomédicale s'inquiète de la « crise de la reproductibilité » (Conner 1986; Dumas-Mallet et al. 2016; Ioannidis 2005b) et la presse en parle aussi (The Economist 2012, 2013a). Nous avons donc analysé si ces observations avaient affecté la façon dont les résultats étaient présentés dans la presse. Nous avons donc comparé le contenu des articles de presse concernant les 19 études publiées avant 2000 et celui des 21 publiées en et après 2000. Les pourcentages d'études invalidées par les méta-analyses correspondantes sont similaires entre les deux périodes (73,4% et 61,9%). Il n'y a pas de différence quant à la description des études comme initiales (52,5% contre 58,8%). Le nombre d'affirmations concernant la robustesse des résultats augmente après 2000 (18,3% contre 27,1%) mais cette différence n'est pas significative. Par contre nous constatons une augmentation forte de la proportion de titres exagérés (test du Chi² : $\chi^2=29,9$, $p<0,0001$) alors que les mentions d'une nécessaire réplification des résultats tendent à disparaître (test du Chi² : $\chi^2=18,5$, $p<0,0001$). Comme ce sont les auteurs des publications qui sont à l'origine de ces mentions, nous avons contrôlé la présence des citations des auteurs dans les articles de presse. Les citations ne diminuent pas et même augmentent. Ces résultats sont présentés sur la figure 4.

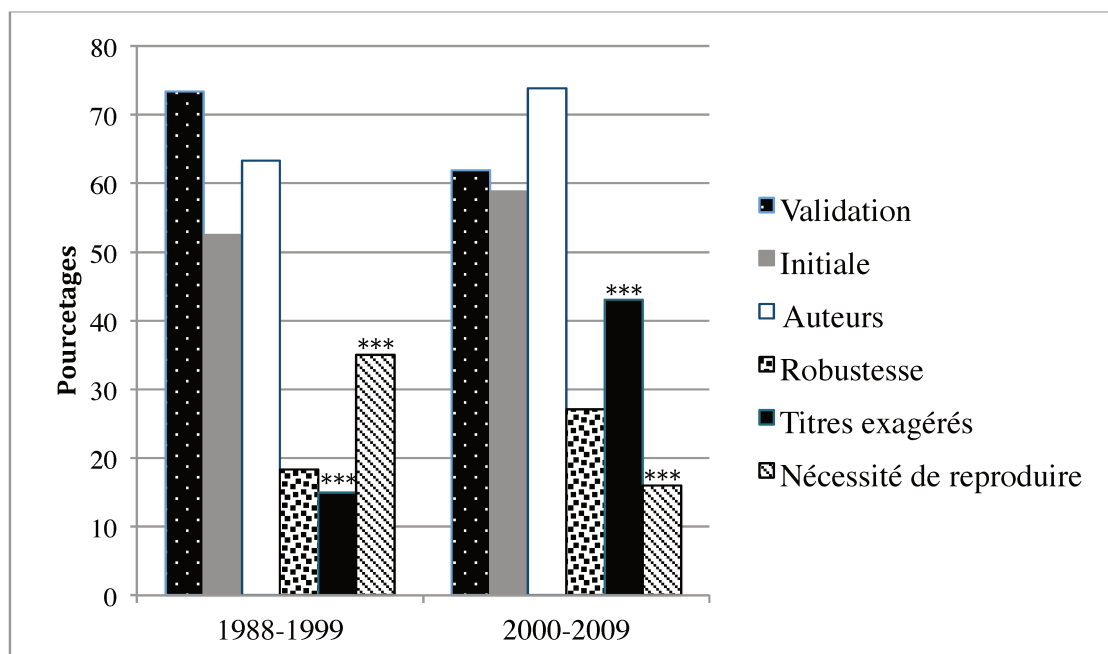


Figure 4 : répartition des différents critères analysés en fonction de deux périodes de publication : de 1988 à 1999 et de 2000 à 2009.

2.3- Apports et limites de l'analyse de contenu des articles de presse couvrant des études initiales

2.3.1- *L'incertitude scientifique dans la presse*

La majorité des résultats initiaux de la recherche biomédicale sont invalidés par les études ultérieures (Dumas-Mallet et al. 2016; Ioannidis 2005b). En général, les auteurs des publications scientifiques mentionnent, à la fin de la discussion, que les résultats sont à confirmer et que la recherche est en cours. Ils ajoutent aussi que d'autres études, sur de plus grandes populations sont nécessaires pour valider les résultats (Angell et al. 1994). Ces limitations sont en général présentes dans les publications scientifiques. Les journalistes, qui couvrent préférentiellement les résultats d'études initiales (Dumas-Mallet et al. 2017; Gonon et al. 2012), devraient donc mentionner que leur validité est soumise à leur reproductibilité. Ainsi, une étude s'est intéressée à la quantification des mentions de nécessité de reproduire les résultats (Holtzman et al. 2005). Les auteurs ont analysé le contenu de 228 articles de presse et programmes télévisuels. L'importance de la réplication n'était mentionnée que dans un tiers des cas (Holtzman et al. 2005). Dans une étude précédente, ces mêmes auteurs avaient interviewé 16 journalistes (Mountcastle-Shah et al. 2003). Pour 79% d'entre eux, il est essentiel de mentionner la nécessité de reproduire les résultats (Mountcastle-Shah et al. 2003). Les résultats de notre étude sont en accord avec ces observations. La mention de la nécessité de reproduire les résultats dans les 426 articles de presse couvrant des études initiales est même un peu moins fréquente (21%). D'autre part, ces mentions sont le fait des scientifiques. En effet, la moitié est faite par les auteurs de l'étude et un tiers par des experts scientifiques du domaine. Curieusement, ces mentions sont moins fréquentes lorsque les études couvertes ont été validées par la suite. On peut donc supposer que les scientifiques interviewés ont une certaine perception de la fiabilité intrinsèque d'une étude initiale et qu'ils sont plus enclins à en faire part au journaliste lorsqu'ils ont des doutes. En tous cas, cette observation suggère fortement que les chercheurs sont presque toujours à l'origine de la mention de la nécessité d'une réplication. Il s'en suit, qu'*a contrario*, les journalistes ne prennent pas l'initiative de mentionner la nécessité d'une réplication, même s'ils semblent avoir compris qu'il s'agit d'une étude initiale. Ceci est en accord avec les études qui montrent que les journalistes considèrent les journaux scientifiques et notamment les journaux prestigieux comme des sources fiables (Hansen 1994; van Trigt et al. 1994; Wilkie 1996).

Notre étude montre qu'il existe une corrélation inverse entre l'exagération du titre et la mention d'une nécessaire réplication. Autrement dit, plus le titre de l'article de presse est exagéré et moins il sera fait mention de cette nécessité dans le corps de l'article. Les journalistes et leurs rédacteurs en chef utilisent de tels titres accrocheurs pour attirer

l'attention des lecteurs. Ceci pose donc un problème majeur puisque les lecteurs se font souvent leur première - et parfois seule impression - à la lecture du titre de l'article de presse. Ainsi des titres exagérés leur donnent une image biaisée des résultats de la recherche. Par exemple, des auteurs ont montré que l'exagération des résultats des études génétiques dans la presse a conduit à ce que les lecteurs s'attendent à ce que les articles se faisant l'écho d'études génétiques parlent de traitements (Caulfield 2004; Caulfield et al. 2012). D'autre part, les scientifiques et leurs institutions, en distribuant des communiqués de presse aux journalistes, sont aussi responsables. En effet, les exagérations présentes dans les articles de presse sont déjà souvent présentes dans les communiqués de presse (Sumner et al. 2014; Sumner et al. 2016; Yavchitz et al. 2012). Ainsi, nos résultats suggèrent que lorsque les scientifiques insistent sur l'importance de reproduire les résultats, ils agissent indirectement sur l'exagération du titre : les journalistes et leurs rédacteurs semblent plus prudents.

Dans cette étude, nous avons considéré que le fait de mentionner la nécessité de reproduire les résultats comme un indicateur. La présence de cette mention impliquait que les journalistes avaient compris et transmettaient l'incertitude scientifique liée aux résultats initiaux. Cependant, la couverture médiatique de chacune des 40 études n'était pas homogène. Ainsi pour une même étude couverte par plusieurs articles, certains pouvaient mentionner l'importance de reproduire les résultats et d'autres non. Seules deux des études initiales avaient une couverture homogène sur ce point. Pour les 38 autres, la mention d'une nécessaire réplication pouvait être présente, ou non alors que les mêmes scientifiques étaient cités. Nous proposons plusieurs hypothèses pour expliquer ces différences. La première est que les scientifiques n'ont pas toujours dit la même chose, et qu'ils ont pu mentionner l'importance de reproduire les résultats à certains journalistes et pas à d'autres. Une autre hypothèse pourrait être que cette mention a été effectivement écrite par le journaliste mais vers la fin de son article et que cette mention a été supprimée par la suite par le rédacteur en chef par faute d'espace. Enfin, certains journalistes et leurs éditeurs pourraient tout simplement ne pas vouloir parler de l'incertitude. Ainsi, des auteurs ont souligné que « l'incertitude en elle-même n'intéresse pas les journalistes. Ce qu'ils veulent ce sont des résultats définitifs ou controversés » (Kitzinger et al. 1997). Cependant, les scientifiques sont aussi responsables. En effet, nos résultats montrent que depuis les années 2000, les scientifiques mentionnent de moins en moins cette nécessité de confirmer les résultats.

2.3.2- Limites de l'étude

Nous avons analysé la couverture médiatique de 40 études initiales portant sur des associations entre des facteurs de risque et 12 pathologies. Nous ne pouvons donc pas généraliser ces observations à d'autres pathologies ou aux études cliniques ou précliniques.

D'autre part, ces études ont été identifiées à partir de méta-analyses incluant au moins sept populations. Nous avons défini cette approche ainsi parce que nous voulions pouvoir corrélérer la reproductibilité des études initiales et leur couverture médiatique. Les méta-analyses nous permettent donc d'évaluer la validité de chacune de ces études initiales. Cependant, nous ne pouvons pas confirmer que ces observations sont applicables à des études initiales qui n'ont pas fait l'objet de méta-analyses.

Nous avons choisi d'étudier la couverture médiatique d'études initiales dans la presse plutôt que les programmes télévisuels. En effet, il est souvent difficile d'identifier une étude scientifique lorsqu'elle est évoquée dans les programmes télévisuels (Bourdaa et al. 2015). Il est aussi vraisemblable que les articles de presse écrite influencent les réalisateurs.

2.4- Conclusion

Les journalistes couvrent préférentiellement les résultats d'études initiales qui sont souvent invalidés (Dumas-Mallet et al. 2017; Gonon et al. 2012). Ils choisissent ces études publiées dans des journaux prestigieux et souvent accompagnées de communiqués de presse parce qu'ils considèrent les journaux scientifiques comme des sources fiables et désintéressées (Hansen 1994; Wilkie 1996). Ces études ne sont pourtant pas plus fiables que celles publiées dans des journaux moins prestigieux (Dumas-Mallet et al. 2016) et on peut légitimement douter du caractère désintéressé des journaux scientifiques. En effet, leurs choix sont influencés par l'attrait potentiel pour les médias des études publiées (Bucchi 2015; Chew et al. 2007). Ces observations permettent de comprendre en partie pourquoi les journalistes ne prennent que rarement l'initiative de parler de l'incertitude liée aux résultats initiaux : ils ne semblent pas conscients que la majorité ne sera pas reproduite. Ils présentent donc les avancées scientifiques comme des prouesses et des faits plus certains qu'ils ne le sont. Ils offrent alors une image biaisée des résultats de la recherche biomédicale au public. Des auteurs ont d'ailleurs critiqué vertement cette attitude passive, qu'ils soient eux-mêmes journalistes (Dentzer 2009; Watts 2014) ou chercheurs en sciences de la communication (Nisbet et al. 2002b; Oransky 2015; Schwartz et al. 2004). En particulier, ces auteurs s'inquiètent de la « décadence du journalisme » (Watts 2014) et de la légèreté avec laquelle les journalistes s'appuient sur les communiqués de presse pour écrire leurs articles. Ils critiquent aussi leur façon de couvrir les résultats de la recherche sans essayer de les replacer dans un contexte plus large ou de réaliser une enquête et une analyse critique. Ainsi Oransky parle du « *single study syndrome* » où les journalistes rapportent les résultats d'études initiales sans en préciser les limites (Oransky 2015). Notre étude de la présentation de l'incertitude permet ici d'apporter des observations quantitatives qui éclairent non seulement le rôle des

journalistes mais aussi celui des scientifiques. En effet, nos résultats montrent que lorsque les scientifiques sont cités dans le texte de l'article de presse et qu'ils précisent que les résultats doivent être confirmés –faisant ainsi part de leur incertitude- les titres de presse sont moins souvent exagérés. Malheureusement, nous avons vu que les citations des scientifiques mentionnant l'importance de reproduire les résultats étaient de moins en moins fréquentes alors que paradoxalement, le problème de la « crise de la reproductibilité » est de plus en plus présent dans la littérature biomédicale (Figure 5).

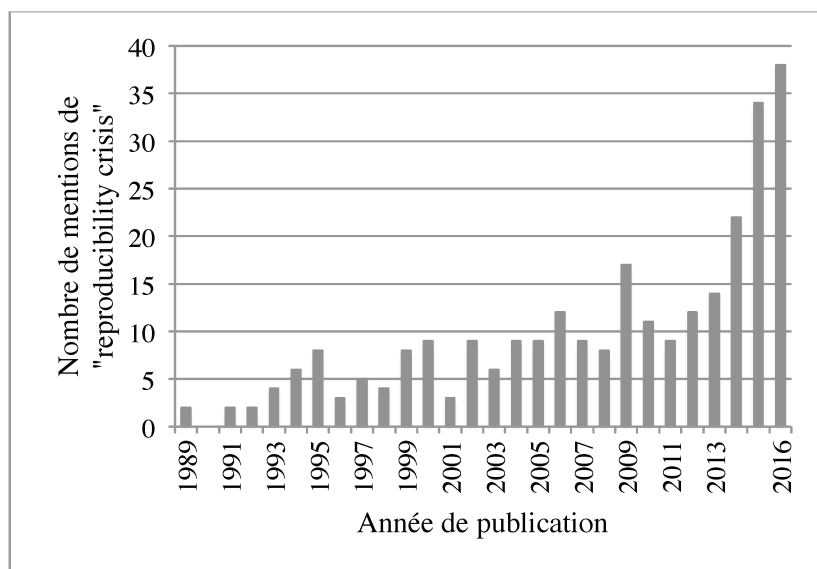


Figure 5 : Graphe représentant le nombre d'articles référencés dans PubMed et mentionnant le terme « *reproducibility crisis* » en fonction des années de publication depuis 1989 jusqu'à 2016

Journalistes et scientifiques sont donc largement responsables de la présentation trompeuse voire fautive des résultats de la recherche biomédicale. C'est le cas pour les scientifiques lorsqu'ils omettent par exemple d'insister auprès des journalistes sur l'importance de reproduire les résultats, ou qu'ils exagèrent les implications de leurs études (Holtzman et al. 2005) ou encore lorsqu'ils « parlent trop » sur des sujets pour lesquels ils ne sont pas compétents (Treise et al. 2002). Quant aux journalistes (et à leurs rédacteurs en chef), c'est le cas lorsqu'ils exagèrent les titres - alors même qu'une étude de 1959 a montré que l'exagération du titre n'avait pas d'influence sur la lecture de l'article ou non (Withey 1959) -. C'est encore le cas lorsqu'ils décrivent les résultats comme des résultats initiaux mais qu'ils ne mentionnent pourtant pas l'importance de les répliquer pour les confirmer -bien que 79% d'entre eux s'accordent à dire qu'il est essentiel de le vérifier - (Mountcastle-Shah et al. 2003). Ainsi, certains auteurs recommandent aux journalistes d'être plus prudents lorsqu'ils rapportent les résultats d'études initiales (Brown 2012; Schwartz et al. 2004).

Ces précautions, contrairement aux idées reçues, sont bien perçues par le public. En effet, des études de réception montrent qu'un résultat de recherche biomédicale est jugé d'autant plus crédible que les scientifiques évoquent leur enthousiasme, mais aussi leurs

doutes (Jensen et al. 2012; Jensen et al. 2013). D'autre part, une étude dans le domaine de la publicité a montré que les experts étaient jugés d'autant plus crédibles qu'ils exprimaient leur incertitude (Karmarkar et al. 2010). Ceci pourrait donc inciter les scientifiques à rester prudents sur la validité de leurs résultats lorsqu'ils décrivent les résultats d'études initiales aux journalistes.

Il n'est resté pas moins que la façon dont les résultats initiaux de la recherche biomédicale sont présentés dans la presse est vraisemblablement préjudiciable puisque le public est mal informé et que ceci risque de miner sa confiance dans la recherche biomédicale et plus largement dans la science. Si, partant de là, on se place dans un registre normatif, il nous semble que ces deux acteurs, scientifiques et journalistes devraient essayer de mieux tenir compte de l'incertitude scientifique : les scientifiques en essayant d'insister sur l'importance de reproduire les résultats et les journalistes en essayant d'inclure ces précautions dans leurs textes.

En tout état de cause, au cours de ce chapitre nous avons identifié et confirmé des facteurs qui sont susceptibles d'attirer l'attention des journalistes. Afin de mieux comprendre les choix opérés par les journalistes lorsqu'ils couvrent les résultats de la recherche, et en particulier de la recherche biomédicale, nous avons réalisé une enquête auprès de 21 journalistes scientifiques. Les résultats de cette étude sont présentés et discutés dans le chapitre suivant.

CHAPITRE IV

Journalistes scientifiques - pratiques et représentations

Décrire, présenter et expliquer les sciences sont les missions du journaliste scientifique. Il se trouve à la croisée de deux disciplines, la science et le journalisme. Il est donc soumis aux normes et contraintes de chacune. Pour les scientifiques, il est important de voir leurs travaux présentés dans les médias. Ils cherchent donc à attirer l'attention des journalistes. Encadrés de manière différenciée par leurs hiérarchies respectives, les journalistes eux choisissent plus ou moins librement les sujets qu'ils vont couvrir. Ainsi dans un article de 1950, D.M. White analyse les raisons qui poussent *Mr Gates* à choisir ou non les sujets à publier dans le journal pour lequel il travaille (White 1950). Les journalistes sont décrits comme des *gatekeepers*, des « portiers » qui contrôlent l'accès ou non de certaines informations à la sphère publique. Ils seraient donc le point de passage, le filtre obligé pour qu'un fait, un événement, une découverte apparaisse dans les journaux. C'est à partir de cette définition des journalistes en tant que *gatekeepers* entre la science et le grand public que nous examinerons leur rôle dans la présentation des sciences dans les médias et plus particulièrement la presse écrite.

En effet, les canaux du journalisme traditionnel restent le moyen privilégié par le public pour obtenir des informations sur les découvertes et les progrès scientifiques. Dans un récent ouvrage des *National Academies of Sciences, Engineering, and Medicine*, les auteurs constatent que la majorité des américains se tourne vers la presse, spécialisée ou non, sous ses formats papier ou en ligne pour obtenir des informations en sciences et technologies (National Academies of Sciences 2017).

L'étude des journalistes, de leurs rédacteurs en chef et des journaux pour lesquels ils travaillent est donc clé pour comprendre la communication des sciences au grand public. Dans ce chapitre, nous décrivons les études et enquêtes qui ont été déjà réalisées auprès des journalistes scientifiques. Nous introduisons ensuite la théorie du *gatekeeping* et expliquons les tensions et difficultés qui existent, pour le journaliste scientifique, à se positionner en tant que « portier » lorsqu'il rapporte les résultats de la recherche scientifique. Enfin, nous décrivons et explicitons certaines de ces tensions à la lumière d'une enquête que nous avons réalisée auprès de 21 journalistes scientifiques.

I- Journalisme scientifique : approche sociologique

1.1- Caractéristiques du journaliste scientifique

Appliquer l'étiquette de profession au journalisme est discutable, comme le discute Erik Neveu dans son ouvrage sur la sociologie du journalisme (Neveu 2009). Les journalistes sont en effet ainsi définis lorsqu'ils possèdent une carte de presse mais son obtention n'est conditionnée par la possession d'aucun diplôme. Même si les journalistes tendent fortement à partager une éthique et des valeurs communes - telles que celles de l'objectivité du travail d'enquête - et qu'ils se définissent comme des observateurs neutres - dont le rôle est de collecter et de transmettre les informations sans influence extérieure - (Shoemaker et al. 1996), il n'existe aucun prérequis dans leur formation. C'est ainsi que la plupart des journalistes apprennent « sur le tas ».

C'est encore sans doute encore plus vrai pour les journalistes scientifiques. Il n'y a pas de consensus sur la nécessité ou non de détenir un diplôme scientifique puisque, aux yeux des journalistes généralistes, ce qui importe ce sont les qualités journalistiques de curiosité et d'investigation. Elles sont primordiales pour expliquer clairement les sujets complexes à des publics profanes (Marchetti 2002). Les avis sont cependant partagés selon que l'on interroge des journalistes scientifiques avec une formation en sciences biologiques et médicales ou non. Les premiers considèrent que leur formation leur permettra d'être moins déférents vis à vis de la science et des résultats scientifiques, qu'ils poseront des questions plus pertinentes. Les seconds pensent qu'être spécialiste dans un domaine des sciences n'est d'aucune aide lorsque le journaliste doit couvrir d'autres domaines que ceux pour lesquels il est spécialisé et que cette spécialisation peut même induire les journalistes à oublier leur esprit critique (Nelkin 1995). On voit donc déjà que le degré de formation d'un journaliste pour qu'il soit considéré ou se considère comme un journaliste scientifique ne fait pas consensus.

On peut aussi se demander si le statut de journaliste scientifique implique des pratiques différentes de celles des autres journalistes. Ainsi, pour Tristani-Potteaux, les journalistes scientifiques « sont des journalistes comme les autres » (Tristani-Potteaux 1997). Selon elle, les reproches qui leur sont faits de trop grande proximité et connivence avec leurs sources ou de leur manque de distanciation et de vision critique face à leurs sujets sont caduques. Pourtant plusieurs études basées sur les entretiens avec des journalistes scientifiques ont montré qu'ils avaient des difficultés à remettre en cause la parole des scientifiques (Logan et al. 2000) et qu'ils leur portaient même une confiance trop importante comme le suggère cette citation d'un rédacteur scientifique britannique : « *si un type a publié dans Nature et que le papier a été évalué par les pairs je n'ai aucun problème à citer tout ce*

qu'il dit sans sourciller » (Hansen 1994). John Lear, rédacteur en chef du *Saturday review*, critiquait aussi cette trop grande proximité avec les sources « *l'esprit d'enquête et le scepticisme, indispensables dans toutes les autres rubriques, sont aussi impératifs pour le journalisme scientifique* » (Nelkin 1995). On voit donc que les opinions divergent dans la littérature et que les journalistes scientifiques ne sont peut-être pas exactement des journalistes comme les autres. Dominique Marchetti les présente également comme un groupe un peu à part pour lequel il existe une certaine entraide et une solidarité (Hansen 1994; Marchetti 2010). Ils auraient ainsi des valeurs un peu différentes de celles des autres journalistes privilégiant la rédaction de dossiers plutôt que la couverture des *hard news*²².

Leurs conditions de travail sont aussi plus difficiles. En effet, dans la plupart des journaux généraux la rubrique « Science » a une position délicate comme l'illustre cette citation d'un rédacteur en chef de la rubrique politique : « *Les sciences gouvernent notre vie, mais la politique (ou ce que l'on considère comme politique) est à la Une des journaux (...). Donc, la science se retrouve dans les premières pages ou à la Une uniquement quand un politicien parle des sciences* » (Badenschier et al. 2012).

Le manque d'espace dans les journaux empêche aussi une couverture adéquate des sciences. Comme le déplorait le scientifique Carl Sagan, « *presque tous les journaux américains ont une rubrique quotidienne dédiée à l'astrologie. La majorité n'a même pas de rubrique science hebdomadaire* » (Hartz et al. 1997). Si les nouvelles et les découvertes récentes trouvent facilement leur place, les articles de fond voient eux leur publication restreinte (Tristani-Potteaux 1997). La dimension d'*infotainment*, c'est à dire informer en distrayant, est souvent prégnante dans la couverture des sciences.

D'ailleurs, ce n'est pas un hasard si la rubrique « Science » est souvent placée au même niveau que les rubriques « Sports » et « Faits Divers » dans les médias populaires : elle est à la fois peu considérée dans la hiérarchie interne du journal mais est en même temps très lue (Neveu 2009). Ainsi, dans les grands organismes de presse, la couverture des sciences sera assignée à un journaliste dédié à cette rubrique tandis que pour des organisations de petite taille, elle sera assurée par un journaliste généraliste qui utilisera plus volontiers les données fournies par les agences de presse (Weigold 2001). C'est la première rubrique à être réduite en cas de scoops (Friedman et al. 1986) ou lors de restructurations. Ainsi, en 1989, 95 journaux américains avaient une rubrique « Science » hebdomadaire contre 19 en 2012

²² *Hard news* : information qui renvoie aux registres de l'imprévu (catastrophe), de l'événement (grande compétition sportive), d'une actualité chaude tant par son immédiateté que par ses enjeux (prise d'otages, débat législatif) Neveu 2009 p 7.

(National Academies of Sciences 2017). On voit bien la dualité qui existe entre l'intérêt du public pour les sciences et son traitement dans la presse.

Or, si les sciences font rarement la Une dans le cas des découvertes spectaculaires, les cas de fraudes ou de scandales trouvent plus facilement leur place dans les premières pages. Dans le cas des « affaires », la presse s'inscrit alors en contre pouvoir. Les articles sortent alors de la rubrique « Science » pour arriver dans les pages « Politique », « Economie » ou « Société » (Champagne et al. 1994; Hansen 1994; Marchetti 2010). Elles ne sont plus traitées par des journalistes scientifiques mais par des journalistes politiques ou des journalistes généralistes (National Academies of Sciences 2017) qui n'ont ni les mêmes formations ni les mêmes compétences (Tristani-Potteaux 2001).

Malgré ces difficultés, les journalistes qui se sont intéressés aux sciences, par choix ou au hasard des restructurations au sein de leur journal, sont attachés à cette rubrique et n'envisagent pas d'en changer (Marchetti 2002). Certains se disent même prêts à renoncer au journalisme pour pouvoir continuer à travailler dans le domaine de la science et de la recherche (Tristani-Potteaux 1997). Ils forment aussi une population à part des autres journalistes : ils sont moins nombreux que les journalistes d'autres rubriques. Ainsi, on recense 3100 journalistes sportifs inscrits à l'UJSF (Union des Journalistes de Sport de France - <https://www.ujsf.fr>). Pour les journalistes scientifiques inscrits à l'AJSPI (Association des Journalistes Scientifiques de la Presse d'Information), on recense 240 inscrits, soit 13 fois moins (<https://www.ajspi.com/fr>). La proportion de pigistes est elle aussi bien supérieure à celle observée chez les autres journalistes. Ainsi, plus de 50% des inscrits en 2012 sur le site AJSPI sont des pigistes, contre 21% pour les journalistes toutes rubriques confondues (Neveu 2009). Par ailleurs, nous savons que les journalistes scientifiques sont en général plus diplômés (Hansen 1994) : en effet 96% des journalistes scientifiques interrogés en 1997 ont un diplôme d'études supérieurs contre 68,8% tous journalistes confondus. Par contre, ils ont rarement suivi une formation en journalisme : seuls 31% sont diplômés d'une école de journalisme (Tristani-Potteaux 1997).

La population des journalistes scientifiques est donc hétérogène. Elle est relativement peu nombreuse et peu de journalistes sont salariés au sein d'un média. Ceci explique sans doute le traitement de qualité inégale des informations scientifiques en fonction du journal qui les couvre et des ressources que celui-ci attribue à la couverture des sciences (Hansen 1994). Les journalistes scientifiques sont également hétérogènes dans leurs diplômes et dans leur définition des critères qui font un « bon » journaliste (Hansen 1994). Si certains privilégient

l'esprit critique et la curiosité, atouts indispensables du journaliste, et laissent entendre qu'une spécialisation en sciences n'est d'aucune utilité si le domaine à couvrir est différent, les autres pensent au contraire qu'une connaissance des sciences est un atout indéniable pour être capable de les expliciter au grand public et que les compétences journalistiques s'acquièrent avec le temps. Malgré ces différences, les journalistes scientifiques sont plus solidaires entre eux que les journalistes des autres rubriques, surtout dans la mesure où les scoops en sciences sont rares. En effet, les conférences et symposiums sont affichés, les journaux scientifiques, les institutions pour lesquelles les scientifiques travaillent préparent de nombreux communiqués de presse, pour certains soumis à embargo et la loi Ingelfinger²³ empêche les chercheurs de parler de leurs résultats à la presse avant que ceux-ci ne soient publiés dans des journaux à comité de lecture. Enfin, dès que la science « sort » de sa rubrique, dès que les sujets abordés ont une dimension politique, économique ou judiciaire, le traitement de cette information échappe aux journalistes scientifiques, pourtant les plus aptes et les mieux formés pour la traiter, et sont alors exposés dans d'autres rubriques et selon des angles²⁴ différents (Champagne et al. 1994; Marchetti 2010).

1.2- Les normes journalistiques

Comme nous l'avons esquissé en introduction, scientifiques et journalistes sont soumis à des normes différentes (Valenti 1999). Pour les scientifiques une découverte ne devient avérée que lorsqu'elle a été présentée puis acceptée et vérifiée par des pairs. La confirmation et la validation par les pairs sont donc primordiales. Ainsi, le savoir scientifique se construit lentement par accumulation de résultats pour arriver à un consensus. La présentation d'une nouvelle découverte est toujours incertaine, ouverte à la critique et à la reproduction par d'autres chercheurs. Sa description dans les journaux scientifiques est très précise et normée. Les scientifiques mentionnent les limites de leur travail et présentent leurs conclusions au conditionnel. Ces normes de prudence scientifique sont aux antipodes des critères d'excellence journalistique : une nouvelle doit être récente, spectaculaire et plus ou moins définitive (Geller et al. 2005). Attendre confirmation des résultats par d'autres chercheurs rend la découverte obsolète. Tous les qualificatifs de précaution utilisés par les scientifiques sont à l'opposé de l'écriture journalistique : un fait devient une nouvelle lorsque c'est la première fois qu'on en parle, lorsque le sujet est « brûlant », sensationnel et qu'il a des

²³ Ingelfinger : En 1968, Franz J. Ingelfinger, alors éditeur du *New England Journal of Medicine*, déclara qu'il n'accepterait pas la publication d'articles scientifiques qui auraient été publiés dans un autre journal ou dans la presse au préalable. Cette règle a été suivie par la plupart des journaux scientifiques. (Nelkin 1995)

²⁴ Selon le jargon journalistique, un « angle » est une manière d'aborder un sujet, d'en valoriser une dimension spécifique.

implications importantes et évidentes sur la population (en l'occurrence les lecteurs). Par contre, la confirmation d'une découverte scientifique ne représente pas une priorité pour les journalistes. Les journalistes sont également soumis à des contraintes de temps (Reed 2001). Peu importe la complexité du sujet à présenter, ou la disponibilité des chercheurs pour leur en parler, leur papier doit être prêt à l'heure dite. Les normes scientifiques et journalistiques divergent donc en raison de leur rapport au temps et de leur position face à l'incertitude.

La plupart des journalistes s'accordent sur les critères qui font d'un fait scientifique un fait intéressant pour les médias : il doit être nouveau et avoir des implications pour le lecteur. Les journalistes scientifiques, en sélectionnant les sujets d'une part, et dans la façon dont ils choisissent de les présenter d'autre part, revendiquent ce statut de *gatekeeper*. Leur sélection est largement influencée par la possibilité de transformer un fait en événement. Ainsi les sciences sont souvent présentées dans leurs articles comme une course entre les scientifiques, où ils ont « les premiers à avoir réussi », « les pionniers ». Les journalistes insistent aussi sur le côté sensationnel des découvertes, chaque découverte est « unique », « révolutionnaire », une « avancée sans précédent ». Ils préfèrent donc couvrir les implications de la recherche plutôt que ses résultats (Hansen 1994; Nelkin 1995). Ils ont aussi tendance à interpréter les résultats scientifiques pour les rendre plus intéressants, pertinents et compréhensibles pour le public (Friedman et al. 1999). L'implication directe des découvertes pour les lecteurs est donc toujours recherchée et mise en avant (Hansen 1994; Hodgetts et al. 2008; Wilkie 1996). Comme le dit cyniquement Flaysakier, pour les médias « deux cas représentent une épidémie et trois morts un génocide » (Flaysakier 1997). Ainsi la couverture des sciences dans la presse généraliste est souvent synonyme de découvertes biomédicales et de santé publique (Badenschier et al. 2012; Hansen 1994; Stryker 2002).

Les journalistes scientifiques se veulent des enquêteurs attachés à la précision des faits et à la force de la preuve (Hansen 2016). Seulement, leur niveau de preuve est bien différent de celui des scientifiques : pour ces derniers un fait avéré a été accepté et vérifié par les pairs. Par contre, les journalistes, à cause des contraintes de temps, considèrent souvent que la publication d'une étude dans un journal scientifique prestigieux ou de la parole d'un expert reconnu représentent des preuves suffisantes (Hartz et al. 1997). Ces pratiques sont renforcées, selon les journalistes, par leur impartialité et l'importance qu'ils accordent à l'objectivité. L'objectivité, pour le journaliste, consiste à donner voix à tous les points de vue, quelle que soit l'origine du point de vue. Les journalistes se doivent d'être neutres, pour cela ils donnent une place égale à toutes les opinions même si ils doutent de leur validité (Friedman et al. 1999). Cependant, cette objectivité si chère aux journalistes est un handicap

majeur pour la couverture des sciences : tout le monde n'a pas autorité en science et un fait scientifique ne peut pas être discuté sans support empirique et hypothèses de vérification (Nelkin 1996).

Hartz et Chappell dans leur ouvrage de 1997 ont ainsi comparé les méthodes scientifiques et à celles des journalistes (Hartz et al. 1997). La méthode scientifique part de l'observation, puis viennent les hypothèses. Celles-ci sont alors testées et une théorie est avancée. La théorie est à son tour testée, proposée et examinée par les pairs, puis acceptée ou non pour publication. Les scientifiques disséminent ici un savoir en construction. Les pairs doivent ensuite valider les résultats pour que la théorie avancée devienne preuve. Cette méthode est donc fortement associée à l'incertitude. La méthode journalistique commence elle par l'identification d'un événement (comme une publication scientifique, un congrès) la vérification des faits auprès des sources, l'obtention de commentaires et d'avis contradictoires puis la rédaction de l'article de presse et son acceptation finale par l'éditeur. Le but est avant tout d'être lu (Laurance 1998). La validité des résultats est secondaire.

En pratique, l'excellence journalistique n'est plus uniquement soumise aux normes généralement acceptées d'impartialité, d'objectivité et d'enquête. En effet, le contexte économique et les difficultés de la presse modifient la façon dont les sciences sont présentées : il faut d'abord faire de l'audience, il faut vendre, savoir transformer une actualité somme toute banale en un événement émotionnel et sensationnel. Paradoxalement, les sciences sont de plus en plus présentes dans les médias alors qu'il y a de moins en moins de journalistes pour les traiter (Autzen 2014; Bauer 2012). Le journalisme semble évoluer vers un journalisme de communication où le bon journaliste n'est plus celui qui découvre et enquête, mais celui qui retient l'attention (Neveu 2009). Ces nouvelles normes journalistiques sont assez bien résumées par la phrase utilisée dans les salles de rédaction britanniques : « *Never let the facts kill a good story* » (Neveu 2009). Le journaliste scientifique doit donc posséder deux types de capitaux, un capital de crédibilité lié aux normes journalistiques et un capital économique par sa capacité à conquérir et fidéliser de nouveaux publics (Marchetti 2010).

1.3- *Gatekeeping Theory* : Facteurs influençant le choix des sujets à couvrir

Si la vision théorique des journalistes comme *gatekeepers* est effective, il est alors important de bien comprendre comment ils choisissent leurs sujets, quels sont les facteurs qu'ils prennent en considération, quels sont les impératifs et à quelles les contraintes ils sont soumis. En effet, le *gatekeeping* est « le processus par lequel des nouvelles potentielles sont triées, formulées et mises en avant pour former les nouvelles au sens médiatique ». Ce processus est souvent défini comme une série de décisions prises de publier ou non un sujet à

chacune des différentes étapes : source, journaliste et rédacteur en chef. Le « *Gatekeeping* n'inclut pas seulement la sélection des sujets mais également la façon dont ils sont formulés et gérés » (Shoemaker et al. 2001). Les facteurs qui influencent et conditionnent le choix des événements dignes de l'attention dans les médias ont été décrits dans l'ouvrage de Shoemaker et Reese de 1996 (Shoemaker et al. 1996). Les auteurs distinguent 5 facteurs qui influencent la sélection des sujets. Ces 5 niveaux d'influence ont été étudiés pour des journalistes scientifiques de la presse allemande (Guenther et al. 2013). Ils sont représentés sur la figure 1.

Les 3 premiers facteurs ont trait aux logiques internes et organisationnelles du champ journalistique : le premier niveau de sélection vient des caractéristiques individuelles (personnelles et professionnelles) du journaliste, puis viennent les contraintes liées aux pratiques, aux routines du travail de communication, et enfin les contraintes liées à l'entreprise de communication pour laquelle le journaliste travaille. Le quatrième niveau de sélection est celui imposé par les facteurs externes à l'organisation médiatique. Enfin le dernier niveau de sélection s'exerce au niveau sociétal. Il dépend du système culturel et politique dans lequel le journaliste évolue. On voit dès lors que le journaliste est loin d'être seul à décider et que sa sélection du sujet à traiter est dépendante de facteurs qu'il ne peut pas tous contrôler.

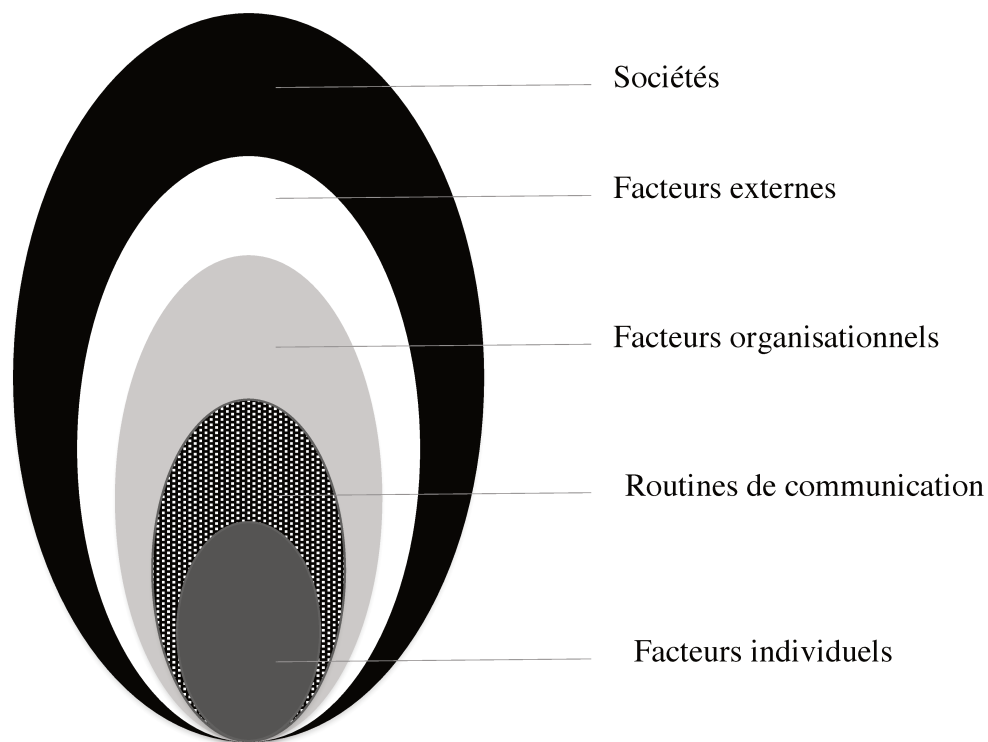


Figure 1 : Les 5 forces influençant la sélection des sujets par les journalistes. Adapté de Guenther et al (Guenther et al. 2013).

1.3.1- Facteurs individuels

Selon les auteurs, les caractéristiques individuelles des journalistes entrent en compte dans leur sélection des sujets (Guenther et al. 2013; Shoemaker et al. 1996). Interviennent tout

d'abord les caractéristiques personnelles telles que l'âge, le sexe, le niveau d'étude. Ainsi il y aurait des sujets typiquement plus masculins comme l'espace, le nucléaire ou l'aéronautique et d'autres plus féminins comme la santé, les sciences humaines et sociales (Tristani-Potteaux 1997). Pour les journalistes scientifiques, leur formation joue aussi un rôle puisque leurs compétences dans un domaine scientifique influence la sélection d'un sujet. L'intérêt que le journaliste porte à certains sujets ou encore son ambition de faire la Une du journal sont autant de critères individuels qui jouent sur la sélection des sujets (Jensen 2010). D'autre part, les valeurs auxquelles adhèrent chaque journaliste et la conception qu'ils se font de leur rôle influencent aussi le choix des sujets. Ainsi, les journalistes remplissent trois rôles différents (Guenther et al. 2013) : celui d'intermédiaire, qui fournit des informations au grand public, celui de critique, qui dénonce les manquements et révèle des affaires, et enfin, celui de fournisseur de divertissements ou de services. Ces différentes conceptions influencent le choix des sujets et la manière de les traiter. Par exemple, un journaliste qui privilégie le rôle critique s'emparera plus volontiers des sujets controversés et débattus. L'idée que la presse se fait d'elle-même en tant que quatrième pouvoir, renforce aussi le système de croyances et de valeurs que le journaliste a de son rôle (Jensen 2010).

1.3.2- Routines et pratiques des journalistes

Dans un environnement où les contraintes de temps et d'espace sont majeures, les routines sont importantes pour permettre au système de bien fonctionner. Ainsi les journalistes mettent en place des routines, des automatismes, un savoir tacite qui agissent comme autant de filtres dans la sélection des sujets. Pour faire face à ces contraintes, les journalistes choisissent des sujets simples. Un sujet long et compliqué à expliquer sera donc bien souvent délaissé au profit d'un sujet plus court et facile à comprendre. Notamment en sciences, où il n'y a jamais assez de temps ni d'espace, le journaliste doit bien connaître son sujet et se garder d'utiliser de trop longues définitions ou des mots trop techniques. Il faut utiliser des mots simples, se baser sur les connaissances que les lecteurs sont supposés posséder et ne pas hésiter à utiliser des métaphores (Siegfried 1991). Les journalistes scientifiques, selon leurs propres mots, racontent une histoire : ils essaient de « raccrocher » le résultat scientifique à son implication pour des personnes malades ou non, pour leur famille (Schafer 2011). De la même manière, le style d'écriture en pyramide inversée favorise les sujets simples (Friedman et al. 1999). En effet, dans ce style d'écriture, l'information considérée comme la plus importante est annoncée dès le début, puis viennent d'autres informations en fonction de l'espace disponible. Ces dernières pourront être coupées au dernier moment par manque de place sans pour autant enlever son sens principal à l'article.

Une autre des routines qui permet aux journalistes de sélectionner leurs sujets provient de la définition implicite, mais assez partagée de ce qu'est une actualité et de sa valeur d'information. Ainsi, dans leur analyse des nouvelles publiées dans la rubrique « International », Galtung et Ruge ont identifié 12 variables associées à un événement et qui augmentent la probabilité qu'il donne lieu à une nouvelle (Galtung et al. 1965). Dans le cas du journalisme scientifique, quelques études montrent que le caractère inattendu, extraordinaire, la charge émotionnelle ou la controverse entourant un fait jouent pour beaucoup. Son intérêt pour le grand public (comme par exemple le nombre de personnes concernées), son implication dans la vie quotidienne (comme par exemple les facteurs de prévention comme la pratique du sport) sont eux aussi des critères pertinents (Badenschier et al. 2012; Bartlett et al. 2002; Guenther et al. 2013; Hansen 1994). La proximité géographique intervient également : le journaliste favorisera des découvertes scientifiques réalisées au niveau local ou national (Badenschier et al. 2012).

Les journalistes se reposent aussi beaucoup sur des sources et des canaux qu'ils ont l'habitude d'utiliser, comme les communiqués de presse, les agences de presse ou les sources officielles (Sessions 2003). Cela leur permet d'identifier et de trier rapidement les nouvelles qu'ils veulent couvrir. Cette pratique est d'ailleurs largement exploitée par les services de communication des différentes institutions scientifiques pour favoriser la couverture médiatique de travaux réalisés en leur sein en fournissant aux médias des textes pré-à-publier (voir paragraphe 1.4 de ce chapitre).

1.3.3- Facteurs organisationnels

Au sein des médias, plusieurs facteurs influencent le choix des sujets. Ainsi les conférences de rédaction - en déterminant les thèmes à couvrir -, le découpage en rubriques - en favorisant certains sujets - et la ligne éditoriale - en privilégiant l'exactitude ou le sensationnel -, jouent tous un rôle. A ceci s'ajoute le processus de « circulation circulaire » de l'information qui fait qu'une information est d'autant plus importante qu'elle a déjà été couverte par d'autres médias (Neveu 2009). La taille de l'organisme de presse, le type de publication (généraliste ou spécialisée), sa fréquence de publication, sa distribution constituent encore d'autres contraintes (Friedman et al. 1986).

La hiérarchie au sein de l'organisation joue aussi un rôle : en dernière instance c'est le rédacteur en chef qui valide les sujets à publier et contrôle leur rédaction. Le journaliste choisit et écrit donc ses sujets en fonction de la connaissance qu'il a de son rédacteur en chef et des attentes qu'il lui prête (Cooper et al. 2002). Des études ont en effet montré que les rédacteurs en chef avaient souvent des critères de définition de la valeur ou de l'importance d'une information différents de ceux des journalistes (Sessions 2003). Dans le cas du

journalisme scientifique, les rédacteurs en chef des rubriques « Science » sont généralement issus des rubriques « Economie » ou « Politique » et n'ont donc pas la familiarité des journalistes scientifiques avec les sciences (Friedman et al. 1986). Ils ont tendance à exagérer le caractère révolutionnaire d'un résultat scientifique (Weigold 2001). Dans une interview auprès de journalistes scientifiques, S. Friedman a relevé les désaccords qui existaient entre eux et leurs rédacteurs. Beaucoup se plaignent de leur méconnaissance des sciences et de leurs nuances (Friedman et al. 1986).

Enfin, le but des entreprises de presse est de vendre. Les contraintes commerciales font que les informations et l'angle selon lequel elles seront traitées sont aussi sélectionnés en fonction des publicitaires et des attentes de l'audience. Comme le souligne une journaliste scientifique d'un journal britannique influent : « *Tous les journalistes ont le même travail : écrire des articles qui vont intéresser nos lecteurs. Les journaux sont une entreprise. Notre travail c'est de faire vendre les journaux. On ne peut pas parler d'un truc juste parce qu'il est digne d'intérêt* » (Jensen 2010).

1.3.4- Facteurs externes

Le quatrième niveau qui influence le choix des sujets, selon Shoemaker et al, est constitué par les facteurs externes à l'organisation médiatique (Shoemaker et al. 1996). Ainsi les groupes d'intérêts sollicitent les médias, envoient des communiqués de presse, organisent des événements. Les publicitaires influencent eux aussi le contenu, en privilégiant des rubriques dont le contenu correspond le mieux à leur produits. Les lecteurs, au travers du courrier des lecteurs, en renouvelant ou non leur abonnement ou en achetant le journal, conditionnent indirectement le choix des sujets.

Enfin les sources jouent un rôle prépondérant, notamment dans le cas du journalisme scientifique. Les chercheurs, leurs institutions scientifiques, les chargés de communication des journaux scientifiques et les services de recherche industrielle produisent régulièrement des communiqués de presse qu'ils envoient aux journalistes. Ceux-ci sont donc soumis à un déluge d'information, qu'ils décident de considérer ou non mais qui déjà leur offre un premier tri des résultats et des événements considérés comme importants par la communauté scientifique.

1.3.5- Facteurs sociétaux

Le dernier facteur qui influence le choix des sujets auprès du journaliste scientifique se situe au niveau sociétal et dépend de la culture et de la politique du pays dans lequel le journaliste travaille. Le modèle des 3 sphères de Hallin explique cette influence de l'acceptabilité sociétale d'une actualité pour que celle-ci soit couverte (Shoemaker et al.

1996). La première sphère est celle du consensus pour laquelle le journaliste ne se ressent pas la nécessité de présenter des opinions contradictoires (Hodgetts et al. 2008) et où son « rôle est de servir d’avocat ou de consécuteur des valeurs du consensus. » La deuxième est la sphère de controverse légitime où les valeurs d’objectivité et d’impartialité sont recherchées : le journaliste présente dans ce cas différents points de vue et cherche des avis contradictoires. Puis vient la sphère de déviance, où le « rôle du journaliste est alors d’exposer, de condamner ou d’exclure de l’agenda public ceux qui s’opposent ou défient le consensus politique. [Cette sphère] marque et défend alors les limites d’un conflit acceptable ».

Cette théorie du gatekeeping nous permet de mettre en lumière les différents facteurs susceptibles d’influencer et même de contraindre les choix des journalistes scientifiques. Ainsi, il apparaît que, si certains sujets peuvent être choisis plus ou moins librement par les journalistes (en fonction de leur intérêt, de leur formation, de leur goût), d’autres leurs sont déjà plus ou moins imposés par les routines qu’ils mettent en place ou l’influence de leur rédacteurs en chef notamment. Dans le paragraphe suivant, nous verrons que les relations que les journalistes entretiennent avec leurs sources, et notamment les chercheurs, jouent aussi un rôle non négligeable dans la sélection des sujets.

1.4- Relations aux sources : interactions entre journalistes scientifiques et chercheurs

Comme nous l’avons déjà signalé plus haut, les chercheurs et leurs organisations respectives ont besoin des journalistes pour promouvoir leur recherche et la rendre visible auprès du grand public. De leur côté, les journalistes ont besoin des scientifiques pour commenter et expliquer les études que ceux-ci ont réalisées. De nombreuses études ont exploré ces relations ainsi que les insuffisances et les inégalités dans la couverture des recherches biomédicales. Ces études ont été publiées aussi bien dans des journaux biomédicaux que dans des revues spécialisées en sciences de la communication ou en sciences humaines (Bubela et al. 2009; Conrad 1999; Cooper et al. 2002; Friedman et al. 1986; Iaboli et al. 2010; Russell 1999). Elles montrent que ces interactions n’ont pas toujours été faciles (Dunwoody et al. 1982) mais ont évolué. Si, comme le soulignait Luc Boltanski et Pascale Madidier en 1970 - « pour les scientifiques la vulgarisation ne peut avoir de définition que négative (...) ne possédant pas de statut nettement défini pour la communauté scientifique » (Boltanski et al. 1970) - de nos jours, les scientifiques recherchent le contact avec les journalistes et la couverture médiatique de leurs travaux. En effet, ils considèrent qu’elle est bénéfique pour leur carrière (Allgaier et al. 2013; Maille et al. 2010; Nelkin 1996;

Peters 1995; Peters et al. 2008; Treise et al. 2002). Les journalistes, quant à eux, se voient comme des intermédiaires entre les scientifiques et le public dont le rôle est de promouvoir et d'engendrer un débat public ; ils opèrent un travail de traduction et ne se considèrent pas comme de simples « vitrines » pour la science (Nelkin 1996; Sicard 1997).

Afin de favoriser et d'améliorer les relations entre les journalistes scientifiques et les chercheurs, des groupes de travail sont maintenant organisés, notamment aux Etats-Unis pour permettre à chacun de comprendre leurs valeurs et contraintes respectives (Schneider 2010). Même si les relations entre les chercheurs et les journalistes sont parfois difficiles, il n'en reste pas moins que leur coopération est jugée comme de plus en plus satisfaisante par les deux parties (Peters et al. 2008; Peters 2012b).

En effet, du fait des contraintes de temps qui pèsent sur les journalistes, il leur est difficile de vérifier indépendamment toutes les études scientifiques qu'ils couvrent. Ils sont donc tributaires de leurs sources pour la vérification des faits qu'ils présentent et s'efforcent donc d'établir des relations de confiance avec elles (Hansen 1994; Hartz et al. 1997; Sicard 1997). Ils se basent aussi beaucoup sur les mêmes sources : les communiqués de presse des grands journaux scientifiques (*Science*, *Nature*, *British Medical Journal* or *The Lancet*) et ceux des institutions scientifiques prestigieuses (Hansen 1994; Kitzinger et al. 1997; van Trigt et al. 1994; Weitkamp 2003; Wilkie 1996). Ils considèrent que ces sources prestigieuses sont essentiellement fiables car soumises à la procédure d'évaluation par les pairs. Ils pensent alors pouvoir se dispenser d'obtenir l'avis de scientifiques indépendants des auteurs de la recherche qu'ils couvrent. Cette (trop) grande confiance dans les scientifiques, en particulier ceux des institutions publiques a été notée par plusieurs auteurs (Jensen 2010; Weigold 2001). Par exemple, un journaliste scientifique du *Philadelphia Inquirer* commente : « *Nous avons tendance à croire ceux qui sont médecins ou ont un doctorat. Nous pensons automatiquement qu'ils savent de quoi ils parlent et nous pensons rarement qu'ils peuvent avoir tort ou agir par intérêt* » (Friedman et al. 1986). Cette trop grande proximité et confiance dans les dires des auteurs scientifiques posent problème car des études ont montré que les découvertes biomédicales publiées par les revues prestigieuses sont souvent invalidées par les observations ultérieures (Dumas-Mallet et al. 2016; Ioannidis 2005b). Les journalistes scientifiques ont donc trop tendance à oublier, dans ce cas, les valeurs journalistiques d'investigation et d'esprit critique.

Les pressions économiques que subissent les journaux grand public font qu'un nombre de plus en plus réduit de journalistes doit couvrir la même quantité de sujets. Ces contraintes de temps croissantes affaiblissent les normes d'exactitude et de vérification. Les journalistes deviennent donc de plus en plus tributaires de leurs sources. De plus, celles-ci utilisent les

critères rédactionnels journalistiques (angle humain mis en avant, écriture en pyramide inversée, sensationnalisme) pour présenter de manière attractive les résultats scientifiques. En effet, on assiste à une professionnalisation des sources : les grandes institutions scientifiques et les revues prestigieuses ne cessent d'étoffer leurs services de communication. Ceci leur permet de promouvoir leurs résultats et activités dans les médias en détournant à leur profit les routines et contraintes des journalistes (Hansen 2016). Ainsi, les attachés de presse, plus nombreux en France que les journalistes eux-mêmes (Neveu 2009) et pour certains issus du journalisme, produisent des communiqués qui pourront s'intégrer facilement dans les pages d'un journal. Les sources officielles sont d'autant plus à même de pousser leurs agendas qu'elles sont déjà en position, aux yeux des journalistes et de leurs rédacteurs en chef, de donner des informations et faits importants (Corbett et al. 1999). Les journalistes ont alors recours à ces sources officielles qui leur fournissent des informations factuelles et qui sont en général disponibles pour répondre à leurs questions (Shoemaker et al. 1996). Souvent l'utilisation régulière d'une même source lui confère un statut de crédibilité et les journalistes se tourneront plus volontiers vers cette source (Dunwoody et al. 1987). Les sources officielles (institutions scientifiques, journaux scientifiques, agences de presse) cherchent à influencer les médias en raison de l'importance qu'elles accordent à leur capacité à attirer l'attention des décideurs et à intervenir dans la mise à l'agenda (Neveu 2009). Ainsi, le problème du journaliste scientifique maintenant n'est pas de trouver des sources, comme le montre S. Friedman dans son analyse de la couverture médiatique de 3 accidents nucléaires (Three Mile Island, Tchernobyl et Fukushima), mais de s'assurer de leur crédibilité (Friedman 2011).

Ces observations sont à rapprocher du concept de *médialisation* décrit dans le chapitre I. En effet, cette professionnalisation des sources et leur utilisation des critères d'intérêt médiatique leur permettent d'influencer le choix des sujets et encore plus dans les conditions actuelles où un plus grand nombre de sujets scientifiques est couvert par un plus faible nombre de journalistes.

D'autre part, même si le journaliste peut effectivement choisir de parler ou non d'un sujet, si il reçoit un communiqué de presse issu d'une revue prestigieuse, qui présente bien le sujet – en respectant les critères d'écriture journalistique –, si c'est un sujet biomédical avec des implications potentielles pour les lecteurs et enfin si le journaliste n'a que peu de temps pour rédiger son article alors cette étude aura beaucoup plus de chances d'être couverte. Ses chances sont d'autant plus grandes qu'elle revêt des qualités que le journaliste ne peut ignorer. Premièrement, ce dernier croit pouvoir se dispenser des vérifications puisque cette étude a été analysée et validée par les pairs et qu'elle a de plus été publiée dans un journal

prestigieux, une source considérée comme fiable par les journalistes. Deuxièmement, si l'étude ne porte pas sur un sujet controversé, elle fait alors partie de la sphère de consensus : le journaliste ne ressent pas le besoin de chercher des opinions contradictoires, il est dans ce cas un pourvoyeur d'information. Il pense pouvoir utiliser les communiqués de presse presque tels quels (Friedman et al. 1986; Samson 2005; Weingart et al. 2016) en ajoutant éventuellement une citation d'un chercheur impliqué dans l'étude. Certains journalistes vont même jusqu'à recopier les communiqués de presse (Veneu et al. 2008). Les journalistes justifient souvent cette pratique minimaliste en considérant qu'elle leur permet de gagner du temps pour traiter de sujets controversés ou qu'ils considèrent comme plus pertinents et nécessitant un réel travail d'enquêteur (Autzen 2014). Lorsque le sujet abordé appartient à la sphère de controverse, le journaliste va alors enquêter et chercher des avis contradictoires notamment dans le cas de scandales comme celui de l'affaire du sang contaminé. Mais la couverture de ce type de sujet échappe progressivement aux journalistes scientifiques pour passer dans les pages « Politique » ou « Economie » et être traitée par d'autres journalistes.

Il y aurait donc deux types de journalisme, comme le souligne Marchetti en parlant du champ médical : un journalisme peu concurrentiel et institutionnel, où le journaliste rend compte des progrès scientifiques et de leurs implications : « l'information de santé se borne, dans ce cas de figure, à un suivi des principales institutions du champ médical ». Il qualifie d'ailleurs ce journalisme de journalisme de routine. L'autre type de journalisme « s'inscrit en rupture relative avec les institutions, la presse prétendant en ce domaine comme dans les autres, être un contre-pouvoir. Elle s'autorise à faire des enquêtes visant à révéler des affaires (enjeu de luttes) ou des scandales et à médiatiser des polémiques sur le mode des débats judiciaires ou politiques ». Dans le cas qui nous intéresse dans cette thèse, il semblerait que les journalistes, rédacteurs des articles de presse que nous avons collectés, « suivent » les communiqués de presse reçus des institutions scientifiques. En effet, ils privilégient les études initiales qui sont pourtant souvent invalidées par les études ultérieures (Dumas-Mallet et al. 2017). Les journalistes n'informent pas le public de ces invalidations, soit parce qu'ils ne les suivent pas soit parce que ces études ultérieures sont publiées dans des revues moins prestigieuses (Chapitre III). Notre analyse du contenu des articles de presse couvrant des études initiales a montré que les journalistes les identifiaient correctement comme initiales dans la majorité des cas. Par contre seulement 21% des articles mentionnaient la nécessité de valider ces résultats par des études complémentaires. D'autre part, dans la majorité des cas, la mention de la nécessité de répliquer les résultats était le fait des scientifiques et non des journalistes (Dumas-Mallet et al. 2017soumis).

On peut donc se demander si finalement, dans ce cas, les journalistes scientifiques remplissent vraiment un rôle de *gatekeeper*. D'autre part, des critiques s'élèvent et dénoncent la dépendance et la trop grande confiance que les journalistes portent aux communiqués de presse et aux agences de presse (Jensen 2010; Lewis et al. 2008; Weitkamp 2014). La journaliste Susan Watts dans les colonnes de la revue *Nature* insiste: « la société a besoin d'un journalisme scientifique capable d'apprécier correctement les valeurs et les vices de la recherche scientifique actuelle » (Watts 2014).

Au total, notre investissement dans la sociologie du journalisme, et notamment notre intérêt pour les questions récurrentes qu'elle soulève, nous a incité à réaliser une enquête auprès de journalistes scientifiques. Sont-ils conscients de l'incertitude inhérente aux résultats d'études initiales? Que disent-ils de leurs pratiques face à cette incertitude? Qu'est-ce qui explique leur discours sur ces thèmes? Voici comment nous avons traduit ces questions sociologiques générales en les adaptant à notre objet de recherche.

II- Une enquête auprès des journalistes scientifiques sur leurs pratiques et leur position face à l'incertitude scientifique

2.1- Présentation de l'enquête

Afin d'éclairer les facteurs influençant les choix des articles scientifiques dignes d'intérêt par les journalistes ainsi que leurs pratiques et les contraintes auxquelles ils sont soumis lorsqu'ils couvrent des découvertes scientifiques, nous avons réalisé une enquête. Vingt et un entretiens semi-directifs ont été réalisés auprès de journalistes scientifiques écrivant pour la presse généraliste ou spécialisée à l'exception d'un journaliste travaillant pour la télévision. Les journalistes ont été choisis pour leur implication et leur intérêt pour les sciences et en tant que rédacteurs d'articles couvrant les sciences et techniques. Les journalistes étaient soit salariés d'un journal, soit pigistes.

Précisions que douze des interviews ont été réalisées par Maud Bernardet qui, après l'obtention de son doctorat en neurosciences s'est intéressée à la médiation des sciences. Sur la base d'une grille d'analyse que je lui ai fournie, elle a donc réalisé les entretiens avec des journalistes scientifiques qu'elle connaissait ou qui lui avaient été indiqués par F. Gonon en fonction de ses contacts personnels.

J'ai sélectionné les neuf autres interviewés et réalisé leur interviews. Quatre de ces journalistes scientifiques ont été interrogés grâce à l'échange chercheur-journaliste proposé chaque année par l'AJSPI et pour lequel j'avais été sélectionnée en 2015 (<https://www.ajspi.com/fr/actualites/lancement-de-la-11e-edition-de-la-bourse-d-echange->

[entre-chercheurs-et-journalistes](#)). J'ai aussi choisi d'interviewer trois journalistes anglais pour pouvoir mettre en parallèle leur discours et les résultats - décrits au chapitre III- après analyse de la couverture des études scientifiques par la presse anglo-saxonne. Enfin, J'ai interviewé deux journalistes de grands quotidiens français, un du *Monde* et un de *La croix*.

D'autres demandes d'entretiens ont été adressées à des journalistes en poste dans d'autres journaux (comme *Le Figaro*, *Libération*, *Médiapart*, *The Guardian*, *The Daily Mail*) mais sans succès malgré plusieurs relances par courrier électronique et téléphone.

Comme le précise le guide d'entretien présenté en annexe 4, quatre grands domaines ont été abordés. Le premier concernait le journaliste et de son champ d'activité, en particulier ses caractéristiques socio-professionnelles (âge, formation, en poste/pigiste, titres pour lesquels il écrit et types de sciences couvertes). Le deuxième domaine portait sur la sélection du sujet (qui choisit et quelles sources de veille sont utilisées), la préparation du document médiatique (temps imparti, vérification des sources, interview des chercheurs, relecture des citations, choix du titre et modification du document). Le troisième domaine avait pour but d'évaluer la position du journaliste par rapport à la recherche (opinion sur la validité de la recherche biomédicale, sur leurs rapports avec les scientifiques). Enfin, venait une dernière question ouverte sur des commentaires que le journaliste pourrait vouloir apporter à l'issue de cet entretien.

Précisons pour terminer ce point méthodologique que tous les entretiens ont été enregistrés et retranscrits dans leur totalité. La durée moyenne d'un entretien est de 48 minutes (16 minutes pour l'entretien le plus court et 81 minutes pour le plus long). Les réponses ont été rendues anonymes et les journalistes ont été identifiés par un numéro et par leur statut (P pour pigiste et S pour journaliste salarié d'un journal). L'analyse et le dépouillement des données de cette enquête ont été réalisés sur la base des retranscriptions et de notes prises lors des entretiens.

2.2- Sociographie des journalistes interviewés

Sur les 21 journalistes interviewés, trois sont anglais et 18 français. Leur moyenne d'âge est de 40,5 ans (24 ans pour le plus jeune et 64 ans pour le plus âgé). Ce sont des journalistes expérimentés puisqu'ils couvrent les sciences depuis 11 ans en moyenne (de 6 mois à 24 ans d'expérience). La parité est presque respectée avec douze hommes et neuf femmes. Le nombre de pigistes interviewé est moins élevé que celui observé dans la population des journalistes français (50%) et ne représente qu'un tiers de l'échantillon (7 pigistes et 14 salariés). Parmi les journalistes salariés, cinq sont des rédacteurs en chef de rubrique « Science » ; quatre pour des revues spécialisées et un pour un quotidien. La répartition en sous groupes est représentée sur la figure 2.

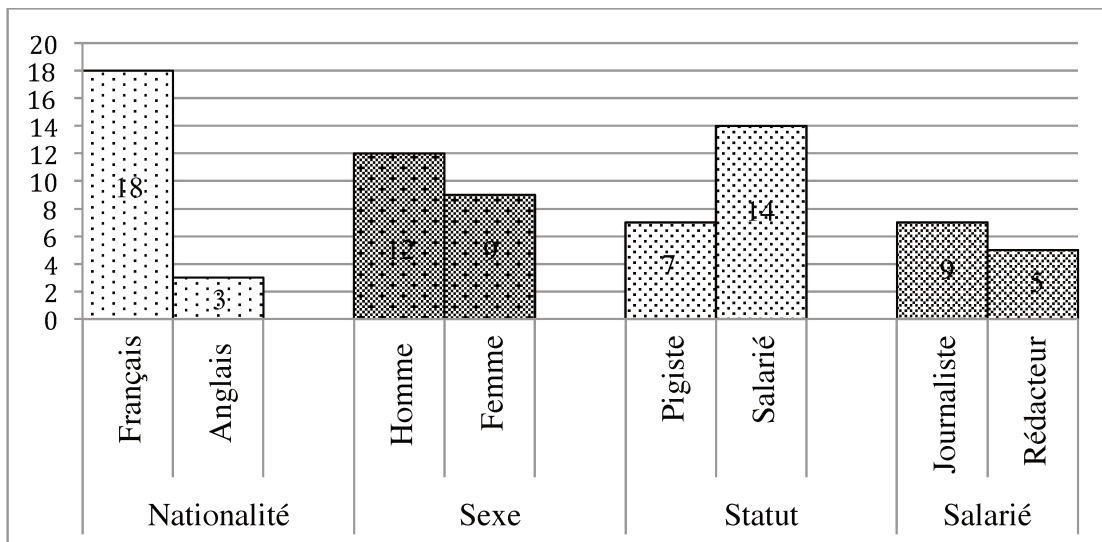


Figure 2 : Répartition des journalistes interviewés en fonction de leur nationalité, sexe et statut.

Tous les journalistes interviewés travaillent pour la presse écrite sauf un qui travaille pour la télévision. Les deux tiers (14 journalistes) écrivent ou travaillent principalement pour la presse spécialisée (comme par exemple *Sciences et Vie*, *Psychologie*, *New scientist*), les sept autres pour des journaux généralistes (*The Daily Mail*, *le Monde*, *La Croix*, *The Times*).

Parmi les journalistes salariés, huit travaillent pour des journaux de presse écrite et 6 pour des journaux ou des sites d'information en ligne (comme par exemple *Lookatscience*, *Atlantico*). Les journalistes de la presse écrite ne fonctionnent pas tous avec les mêmes formats : quatre journalistes travaillent pour des mensuels, deux pour des hebdomadaires, et deux pour des quotidiens.

Pour les pigistes, la majorité écrit uniquement pour des mensuels et des hebdomadaires (5 pigistes). Seuls deux autres écrivent pour des quotidiens mais aussi pour des hebdomadaires.

Les journalistes n'ont pas tous la même formation : cinq ont reçu une formation en sciences uniquement (Physique, biologie, génétique, biochimie), quatre sont issus d'une école de journalisme. Cinq journalistes ont la double casquette : ils sont diplômés en sciences et en journalisme. Enfin, sept journalistes ne sont diplômés ni en sciences, ni en journalisme (études de marketing, langue, Sciences Politiques).

Les journalistes couvrent en général plusieurs domaines des sciences mais avec une prépondérance pour les sciences biomédicales et la santé (37%). Puis viennent la biologie (22,5%) et des sujets de sciences humaines ou des sujets « société » (22,5%). L'écologie et les sciences environnementales ne représentent que 12% des sujets couverts par

ces journalistes. L'astronomie et la physique sont peu suivies par ces journalistes (Figure 3). Près de 60% des sujets couverts ont donc attiré à la biologie et aux sciences biomédicales.

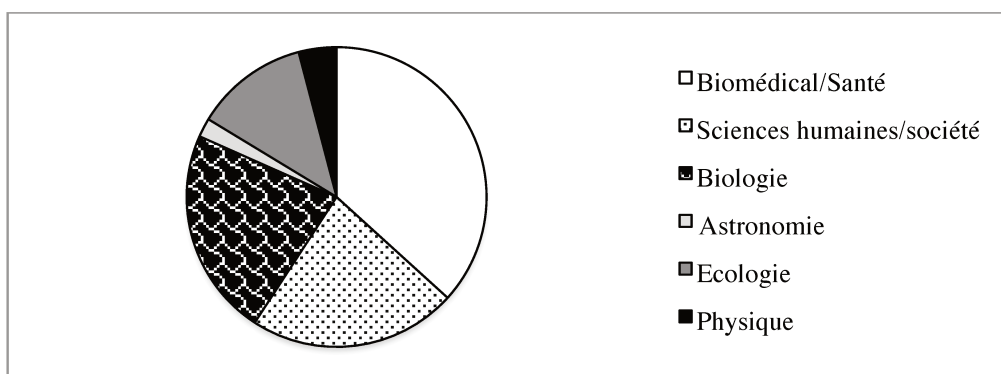


Figure 3 : Répartition des sujets couverts par les journalistes.

2.3- Représentations des choix du sujet et des sources

2.3.1- Choix des sujets

A l'exception des rédacteurs en chef qui décident et valident les sujets, les 16 journalistes salariés et pigistes proposent leur sujets mais répondent aussi à des demandes. Pour les journalistes pigistes, la commande de sujets est aussi fréquente que l'acceptation de sujets qu'ils proposent. Autrement dit, ils choisissent le sujet dans 50% des cas en moyenne. Il existe cependant de grands écarts entre les pigistes puisque certains ne travaillent presque intégralement que sur commande (dans 90% des cas pour deux journalistes) alors que d'autres pigistes à l'inverse, n'écrivent que sur les sujets qu'ils choisissent (dans 90% des cas pour deux pigistes). Pour les trois autres pigistes, la répartition est identique, ils choisissent aussi souvent leurs sujets qu'ils répondent à une commande.

Les journalistes salariés choisissent leurs sujets (dans 63% des cas). Il n'existe que peu de variations entre eux, sauf dans le cas d'une journaliste stagiaire travaillant pour un site d'information en ligne et dont les sujets lui étaient majoritairement imposés. On voit donc qu'en moyenne, il n'y a pas de différence significative entre les pigistes et les salariés quant au pourcentage de sujets choisis et traités. Le statut professionnel n'implique pas, dans le contexte de notre enquête, que les journalistes salariés aient une plus grande liberté dans le choix des sujets qu'ils traitent que les pigistes.

Les journalistes interviewés insistent beaucoup sur les critères qui influencent le choix de leur sujet, sur l'importance du caractère nouveau, sensationnel et de son implication pour le lecteur. Ainsi, un journaliste salarié indique les différents types de sujets qu'il considère comme importants à traiter.

« Il y a des thèmes qui généralement font de bons papiers, qui sont répartis en deux catégories, ou peut être trois. D'abord, ce sont des résultats qui ont une implication

potentielle directe pour les individus, pour les lecteurs, donc oui souvent des choses dans la santé ou la médecine. Mais ça peut-être aussi des trucs en technologie ou autre. N'importe quel sujet qui quand le lecteur va le lire se dira « ben, ça c'est peut-être intéressant pour moi ». La deuxième catégorie c'est ce qu'on appelle des sujets qui font rêver, des trucs sur l'infiniment grand ou l'infiniment petit, des trucs sur les origines de l'univers, les voyages dans l'espace, tous ces sujets à la frontière avec l'imagination, ça ça marche toujours très bien. Et la dernière catégorie c'est les grandes questions du moment, à l'intersection des sciences, de la décision publique, et de la politique etc. Donc qu'on parle du réchauffement climatique, ou de comment le NHS²⁵ pourrait avoir recours à l'ingénierie génétique ou alors de l'utilisation des résultats scientifiques pour prendre des décisions politiques voilà, ces trois catégories c'est celles qui marchent. » S9

Ici, trois critères de choix se dessinent : les sujets qui impliquent personnellement le lecteur, comme par exemple la pratique d'une activité physique pour se prémunir de certaines pathologies, ceux qui le font rêver et le surprennent, et enfin, bien sûr, les sujets qui portent à controverse.

Deux autres journalistes eux ont insisté sur l'importance de la nouveauté et du caractère surprenant du sujet dont ils se font l'écho.

« Enfin, je suis un peu rigide là-dessus et j'ai toujours essayé de trouver des études qui sortaient du lot et qui, vraiment apportaient quelque chose de nouveau et qui ne venaient pas contredire une autre étude. Ou alors qui venaient contredire une autre étude, mais de manière intéressante... » S5

« Ce qui compte c'est est ce que c'est assez nouveau, assez surprenant, est ce que c'est assez intéressant scientifiquement parlant. Et si c'est le cas, alors ça importe peu d'où ça vient du moment que c'est une bonne information, une découverte scientifique qui vaut le coup. » S10

D'autres indiquent que le sujet doit cependant rester simple et compréhensible.

« On reste sur les résultats, on reste sur quelque chose qui est très accessible et lisible très rapidement. Donc on évite le vocabulaire qui serait un tout petit peu compliqué et...voilà. Ca... ça, c'est fait dans le but de toucher un maximum de personnes. » P6

On voit ici que les critères influençant le choix des sujets sont assez partagés par les journalistes que nous avons interrogés et qu'ils sont en accord avec les principaux facteurs décrits dans la littérature (Bartlett et al. 2002; Logan et al. 2000; National Academies of Sciences 2017).

²⁵ NHS : National Health Service. Système de santé publique du Royaume-Uni

2.3.2- Sources utilisées par les journalistes

Au cours de l'enquête, les journalistes ont été interviewés sur les sources qu'ils utilisaient habituellement ou fréquemment comme inspiration pour leurs articles. Les réponses ont été répertoriées dans le tableau 1. Chaque journaliste utilise bien sûr plusieurs sources. Sept sources principales différentes ont été identifiées : (1) la presse (à l'exclusion des journaux scientifiques spécialisés) (2) les communiqués de presse des organismes de recherche (CNRS-INSERM-MRC²⁶), (3) les communiqués de presse des grands journaux scientifiques (*Nature*, *Science*, *Proceedings of the National Academy of Sciences (PNAS)*, *British Medical Journal (BMJ)*, *New England Journal of Medicine*), (4) les sommaires, éditoriaux ou articles des grands journaux scientifiques, (5) les dépêches des agences de presse (*AFP*, *Reuters*, *AP*), (6) les ouvrages spécialisés et enfin (7) le « terrain » par exemple leurs rencontres avec d'autres chercheurs dans des laboratoires dans des congrès, ou des contacts initiés directement par les scientifiques.

Tableau 1 : Sources utilisées par les journalistes pour choisir leurs sujets.

Sources utilisées	Nombre de journalistes
Autres articles de presse	19
Communiqués de presse des organismes de recherche	16
Communiqués de presse des journaux scientifiques	16
Journaux scientifiques (éditoriaux, résumés, articles)	16
Agences de presse	13
Terrain	12
Livres spécialisés	4

Presque tous les journalistes (19/21) lisent la presse dans le cadre de leur veille pour identifier des sujets intéressants à développer. Ils lisent les sites web concurrents, les journaux étrangers (plusieurs journalistes français ont cité *The Guardian* et *The New York Times*), les magazines dédiés aux sciences (*Sciences et Avenir*, *Sciences et Vie*, *New scientist*). Plusieurs ont également cité des blogs ou les réseaux sociaux (Twitter) comme des sources potentielles qu'ils suivent et regardent dans leur veille. La « circulation circulaire » de l'information est donc bien prégnante dans notre échantillon.

La plupart reçoivent les communiqués de presse proposés par les organismes de recherche et par les grands journaux scientifiques comme le souligne un des journalistes :

« *Les journalistes scientifiques sont maintenant bombardés avec des sujets potentiels des journaux scientifiques, des organisations.* » P3

²⁶ MRC : Medical Research Council

Certains les ignorent, parce qu'ils n'y trouvent pas beaucoup d'intérêt et qu'ils reprochent le principe des organisations scientifiques d'envoyer sans distinction les mêmes communiqués de presse à tous les journalistes scientifiques :

« Puis c'est vachement...c'est la même qu'ils vont envoyer à tous les journalistes... les trucs sur la physique quantique, je m'en fous complètement, quoi. En fait, dans le tas d'infos qu'on reçoit que je trouve 1) pas énorme 2) pas si fréquent, y'a beaucoup d'infos dont j'ai strictement rien à faire. Donc au final, je vais m'intéresser à deux-trois trucs. » P1

D'autres critiquent leur mauvaise qualité. Ainsi, un des journalistes salarié a décrié la qualité des communiqués de presse qu'il recevait de l'INSERM.

« Je vais donner l'exemple de l'INSERM. L'INSERM fait des communiqués de presse sur un sujet. J'appelle le chercheur, pour discuter de ce communiqué de presse qui, généralement, me semble aberrant ou... enfin, il y a de grosses contradictions ou des choses qui me semblent sortir de nulle part, ils me disent : « ah, mais comment ils ont pu sortir ça ? » Avec l'INSERM, ça ne loupe pas. » S7

L'enquête montre que si les journalistes reçoivent en effet une grande quantité de communiqués de presse, certains le déplorent et trouvent que leurs contenus sont inutiles, inintéressants et même parfois largement exagérés.

Les journaux scientifiques sont aussi une source importante de sujets (16 journalistes les suivent). Beaucoup ont cité *Nature*, *Science* et *PNAS*. Pour ces journalistes, l'intérêt principal de ces revues scientifiques est que les articles ont été examinés et validés par les pairs. Ils sont pour eux des sources valides et fiables de sujets. Ils les utilisent donc volontiers :

« Nature, donc une revue, quand même...parmi les premières au monde et dont on suppose qu'ils font très très attention, qui avait donc été annoncé à grands fracas et là, je suis un peu tombé dans le panneau. C'est à dire que j'ai quand même pensé, qu'ils ne prendraient pas un tel risque, et qu'ils avaient vraiment vraiment...tout contrôlé. » P6

« Un article scientifique publié dans un journal validé par les pairs, suivant le processus habituel- Ca apporte du poids, ça veut dire que ce n'est pas juste un excentrique qui présente ces résultats. » P3

« Ces gros papiers qui sortent, bien sûr examinés et validés par les pairs, ils présentent des résultats vraiment robustes et fiables. » S9

« Moi je fais confiance au fonctionnement académique parce qu'en l'occurrence on ne peut pas...je ne peux pas juger. » S3

Ces éléments confirment les résultats de la littérature. En effet, si l'on pouvait penser que la citation « *si un type a publié dans Nature et que le papier a été évalué par les pairs je n'ai aucun problème à citer tout ce qu'il dit sans sourciller* » d'un rédacteur scientifique britannique d'un journaliste de 1994 (Hansen 1994) était obsolète, on voit ici qu'il n'en est rien. Malgré le nombre croissant de rétractations dans ces revues, elles restent tout de même toujours considérées par les journalistes scientifiques comme des sources fiables d'informations.

Près des deux tiers des journalistes interviewés utilisent aussi les dépêches des agences de presse. Mais ils expliquent qu'ils regardent plus ce que ces agences proposent qu'ils ne les utilisent vraiment.

« L'AFP était souvent la source première, imprimée, enfin essentielle dans les rédactions. L'AFP c'est vraiment... la dépêche de l'agence, c'est vraiment la base souvent. »
S3

En effet, ils ne leur font pas toujours confiance.

« on a pour politique de vérifier les éléments des dépêches AFP, ça, c'est très original sur le Web, mais on vérifie, parfois, on les réécrit entièrement, on dit juste que notre base était l'AFP, mais on donne des conclusions totalement opposées, parce que quand on regarde, ils sont... maintenant, on a des noms, on sait que si la personne de l'AFP qui a signée, il n'y a pas de nom, on sait qu'il faut absolument relire l'étude de base, parce que c'est... il y a une chance sur deux que ce soit un contresens grossier. » S7

Ils ne les utilisent finalement que dans des cas où les contraintes de temps sont vraiment trop fortes et utilisent ces sources pour des sujets « peu importants » ou « faciles », des brèves ou des « Actu »²⁷.

« Parfois on reprend leur papier (Press Association), vous savez on les republie nous-mêmes. On fait ça, mais c'est plutôt pour les études moins intéressantes qui ne sont pas si importantes. Je pense que quand c'est vraiment une grande découverte, on préfère écrire dessus nous-mêmes. » S10

« On dit « bâtonner de la dépêche ». Bâtonner, ça veut dire on reprend des morceaux d'une dépêche AFP, donc on fait beaucoup confiance à l'AFP. Ce qui est souvent bien, mais des fois mal. Et on meuble, on fortifie, on enrichit ... » S6

Ils légitiment cependant leur pratique à l'égard des communiqués de presse en expliquant qu'en bâtonnant certaines dépêches, ils ont ainsi plus de temps pour réaliser un

²⁷ Les brèves ou « actu » sont des textes courts publiés dans la presse écrite. Elles fournissent des informations sur l'actualité. Elles sont en général produites à partir des dépêches des agences de presse et souvent recopiées ou « bâtonnées »

bon travail d'investigation sur des sujets qu'ils jugent plus importants. Certains journalistes précisent qu'ils utilisent les communiqués de presse pour des brèves, des « actus » qui leur semblent plus triviaux dans leur hiérarchie des sujets à traiter que les dossiers complexes. Ils ne consacrent donc que peu de temps à ces « actus » (moins de 3 heures).

« On fait ça parce que ça nous libère du temps, du temps pour nos journalistes pour écrire de plus gros papiers. » S10

« Le raisonnement derrière tout ça c'est que parce que je vais pouvoir écrire ce papier en 1 heure, j'aurais beaucoup plus de temps pour des trucs beaucoup plus intéressants, où je pourrais vraiment expliquer, et faire un travail d'enquête. » S9

« Donc, les communiqués de presse, c'est vrai que je ne les utilise pas forcément tout de suite pour des gros papiers. C'est plus pour l'actu. » S12

« Quand je bâtonne, quand je suis obligé parce que j'ai pas le temps, parce que je suis pas sur place, je signe pas l'article. Je signe AFP ou AP ou avec agence, ou je sais pas d'autre. » S13

Les journalistes que nous avons interviewés semblent donc utiliser les dépêches d'agence bien qu'ils ne leur fassent pas confiance et qu'ils soient conscients de leurs inexactitudes ou exagérations. Ils s'en défendent et leurs arguments rejoignent ceux décrits dans la littérature : les conditions de travail dans les médias sont difficiles, il y a de plus en plus de sujets à couvrir pour un nombre restreint de journalistes (Autzen 2014). Les journalistes doivent donc faire des choix et utiliser les ressources auxquelles ils ont accès pour écrire certains sujets qui leur semblent « faciles et crédibles » et consacrer plus de temps à des sujets plus compliqués. C'est aussi pourquoi les journalistes insistent aussi beaucoup sur l'importance de combiner plusieurs sources dans la rédaction de leurs papiers. C'est une des bases de leur travail.

« Je n'ai pas pour principe de croire que parce que quelque chose est écrit dans un communiqué de presse vous pouvez juste le répéter. Certains disent « bon c'était dans un communiqué de presse donc ça va, je peux juste le répéter. Ce ne sera pas ma faute, l'erreur ne serait pas la mienne ». Et bien si, l'erreur ce serait la votre. » S9

« C'est toujours pareil, on recoupe. On ne prend jamais une seule source. » S3

« Le principe du boulot, c'est vérifier ses sources et donc savoir, voilà, si c'est vraiment pertinent de parler de cette recherche-là à tel moment. » S14

Il y donc déjà une contradiction entre ce que les journalistes disent et ce qu'ils font puisque si les conditions l'exigent (contraintes de temps), ils recopient les dépêches des agences de presse et que souvent, dans le cas des journaux scientifiques, ils se fient

aveuglément au système de validation par les pairs qui pour eux garantit la fiabilité des résultats publiés. Mais pourtant, ils expliquent que leur expérience et leur connaissance des sources leur permettent de ne faire confiance qu'à bon escient.

« En tant que journaliste vous avez affaire régulièrement à ces agences de presses ou à ces communiqués de presse, vous savez lesquelles croire. Vous connaissez celles qui ont l'habitude d'exagérer les trucs, vous connaissez toutes les organismes de presse, vous savez qui sont les gens ceux qui ont une grande intégrité ceux qui ne feraient pas ça, et ceux qui prendraient des raccourcis et exagérerait les résultats pour faire les gros titres. » S9

Les journalistes utilisent donc principalement comme source les autres titres de presse concurrents ou étrangers. Ils utilisent aussi les communiqués de presse et les journaux scientifiques prestigieux qui pour eux sont des sources valides et légitimes de sujets fiables. Ils peuvent avoir recours aux dépêches des agences de presse par manque de temps pour écrire un papier ou pour pouvoir consacrer plus de temps à des « bons » papiers. Ils insistent également sur l'importance de croiser les sources pour se faire une opinion.

2.4- Préparation du document

La troisième partie de l'enquête portait sur la rédaction des papiers en tant que pratique journalistique. Les journalistes lisaient-ils l'article primaire quand celui-ci était à l'origine d'un de leur sujet ? Quelles sources jugeaient-ils importantes pour la validation de leur sujet ? Contactaient-ils les auteurs de la publication scientifique ou cherchaient-ils à contacter d'autres scientifiques ? De combien de temps disposaient-ils pour écrire leurs articles ? Enfin ils ont été interrogés sur les remaniements qui pouvaient être faits par d'autres sur leurs articles et sur le choix du titre.

2.4.1- Vérification auprès de l'article primaire

Tous les journalistes interviewés affirment lire au moins le résumé de la publication scientifique qu'ils couvrent. Treize journalistes disent lire la publication dans son intégralité. Cependant, cinq de ces journalistes reconnaissent qu'ils ne le font que rarement parce que les publications sont en anglais et que les articles ne sont pas accessibles gratuitement (en particulier pour les pigistes dont les ressources sont moindres si on les compare à celles d'un journaliste de la rubrique science d'un journal). Trois journalistes ne lisent pas de publications primaires parce que leur activité journalistique ne les amène pas à couvrir des résultats d'études scientifiques primaires (un journaliste TV et deux journalistes couvrant surtout des aspects sociétaux). Pour certains des journalistes français, le fait que les articles soient rédigés

en anglais constitue aussi un frein à cette lecture. Les journalistes soulignent aussi le caractère trop technique de certains articles.

« Par contre, je lis pas le corps de l'étude si c'est un... Ah la, mais c'est incompréhensible, la génétique. » S8

« Il y a des articles qui sont assez faciles à lire, donc on lit volontiers, ben que moi, je lis volontiers. Il y a des articles qui sont très techniques en génétique, en biologie moléculaire par exemple, on est très vite perdu. Où il y a beaucoup de technique qui échappe globalement aux journalistes scientifiques, même avec une petite formation scientifique. » S13

Ceci confirme que les articles scientifiques primaires sont des sources potentielles fréquentes de sujets mais que les journalistes n'ont ni le temps ni toujours les capacités de lire les articles primaires dans leur totalité. Ils lisent en général les résumés et les conclusions. Malheureusement, cette limitation de leur lecture constitue déjà un biais puisque les exagérations sont déjà présentes à ces niveaux dans les articles scientifiques (Boutron et al. 2010; Gonon et al. 2011).

2.4.2- Vérification des informations

Selon les normes de sa profession, au moins dans les pays démocratiques, une des missions d'un journaliste scientifique est d'enquêter. Le journaliste est supposé ne pas se contenter pas de recopier ce qui est écrit dans le résumé article scientifique primaire, le communiqué de presse ou la dépêche de l'agence de presse. En accord avec cet idéal, tous les journalistes interviewés insistent sur l'importance de recouper les faits de l'étude avec l'auteur, d'obtenir un commentaire du scientifique.

« L'idée, c'est qu'on part d'une revue primaire, par exemple un papier de Science. On cherche à interviewer le premier auteur ou le directeur de l'équipe. » S12

« On veut toujours quelqu'un pour commenter. Quand j'appelle le scientifique pour avoir un commentaire c'est pour pouvoir le citer, amener un peu de peps à mon article, à le replacer dans son contexte. (...). Je suis réticent, je serais toujours réticent à couvrir une étude si je n'ai pas parlé à quelqu'un qui travaille dans ce domaine et qui connaît le papier et peut me le vendre. Je pense que la plupart des journalistes scientifiques qui se respectent ont la même opinion. » P3

« Ce qu'il y a de mieux quand même c'est d'interviewer la personne qui était en charge de l'étude, ce serait la personne la plus à même de... d'expliquer, de voir quel est vraiment l'intérêt de l'étude, etc. » S5

Les journalistes interviewés ont également souligné l'importance de citer plusieurs sources, notamment des sources ayant une position différente vis à vis des résultats présentés.

« C'est ce qu'on appelle la validation par les pairs pour le journalisme scientifique. On ne fait pas juste confiance à ceux qui ont publié le papier et ses auteurs. On veut que d'autres scientifiques commentent l'étude, juste pour savoir si cette publication est si pertinente que ça, pour savoir ce qu'ils en pensent. » S10

« Il y a le scénario parfait : on parle avec l'auteur, on lit la publication, on lit le communiqué de presse on lit le résumé ; et puis on interroge quelqu'un d'extérieur qui travaille dans ce domaine et qui peut commenter sur l'intérêt de l'étude et puis on peut aussi trouver une troisième personne pour un autre commentaire ou une vision un peu contradictoire, quelqu'un qui n'aurait pas la même opinion sur l'étude. » P3

Mais, en pratique, quand les journalistes rapportent les résultats de publications scientifiques publiés dans des journaux à comité de lecture, ils le font rarement.

« Je vais le chercher [l'avis du chercheur] quand l'information n'est pas claire, ce qui est... ou quand elle... Ils ont... souvent, ils le donnent, leur avis, dans les discussions. Par contre, dès qu'on soupçonne - et, on a beaucoup de raisons de soupçonner - qu'il y a des biais dans l'étude ou qu'il y a des éléments de contexte qui manquent, on va leur demander : est-ce qu'ils pensent qu'il y a des biais. » S7

« Souvent, ce qu'on essaye de faire aussi, même si, à la limite c'est un petit peu heu... on peut dire que c'est superflu, puisqu'il y a le peer-review qui intervient avant la publication de ce genre de papier. Mais souvent, on essaye aussi de passer des coups de téléphone à l'intérieur de la... de cette communauté scientifique, il y a des gens qui n'ont pas participé à l'étude, pour avoir un petit peu une vision de l'extérieur et une perspective qui soit pas issue des chercheurs eux-mêmes qui ont publié le papier quoi. » S13

Là encore, on voit que le travail d'enquête du journaliste est limité lorsque ceux-ci se font l'écho de résultats scientifiques publiés dans des revues prestigieuses. Ceci confirme à nouveau la fiabilité et la validité qu'ils accordent aux articles publiés dans de telles revues et au système du *peer-review*.

2.4.3- Ecriture et remaniements par la rédaction

Les journalistes que nous avons interviewés disposent de délais de rédaction très variables : de quelques heures pour les plus courts jusqu'à six mois pour les plus longs. Pour les temps les plus longs, les journalistes traitent de sujets plus complexes nécessitant de longues enquêtes (comme une enquête sur les médicaments chinois) mais continuent à traiter d'autres sujets en parallèle. En moyenne, les journalistes interviewés disposent de 4 à 30 jours pour écrire leurs sujets. Pour les journalistes salariés, les délais sont beaucoup plus courts que

pour les pigistes - juste quelques heures- alors que les pigistes disposent en général d'au moins 24 heures.

Notre enquête montre que les journalistes scientifiques privilégient eux aussi la rédaction de leur article sur le principe de la pyramide inversée : l'information jugée comme la plus importante est donnée dès le début de l'article. Les précisions et limites sont données ensuite en fonction de l'espace alloué à l'article. Ce style d'écriture leur permet de légitimer le sensationnalisme souvent associé aux informations scientifiques :

« C'est pour ça qu'on écrit nos papiers comme ça : on les écrit pour que même si on coupe les 2/3 à la fin, le papier soit toujours lisible. » P3

Un des journalistes semble d'ailleurs apporter un regard un peu critique sur ce style rédactionnel et sur les conséquences de « l'embellissement » de la nouvelle sur la compréhension et l'information mises à la disposition du lecteur :

« Journalistiquement parlant, on va vous dire « ah non, d'abord, l'info. » Comme si il faut tout résumer sur une info, et puis donc, vous catapultez le résultat et évidemment, vous ne négligez pas donc d'employer, des adjectifs en disant merveilleux, extraordinaire, réussi et tout ça. C'est terrible ça. Parce que du coup, il faut que j'en... ça c'est un peu moins important, j'enlève, j'enlève. Il faut garder le truc. Et donc du coup, faut garder... c'est là qu'on a tendance aussi à embellir la chose, parce qu'on dit attends, je vais enlever ça et ça, j'ai plus que ça. C'est un peu rêche, c'est un peu dur, c'est pas très affriolant et du coup on rajoute un petit adverbe, un petit qualificatif, pour dire merveilleux, formidable, etc. Pour valoriser son... Et on biaise un petit peu quoi, c'est fou. Mais d'un côté, l'homme de la rue, le Monsieur qui est là, il va peut-être lire, 1/4 de page, mais il lira pas 3/4 de pages. » S6

Dans l'ensemble, leurs articles sont peu remaniés par la rédaction, ou juste pour en corriger l'orthographe. Par contre, dans les cas rares où leur document est modifié en profondeur, leur position vis à vis de ces remaniements est très variable. Certains sont mécontents parce que ces modifications engendrent des contresens ou des exagérations du fait de la suppression de certains qualificatifs, du raccourcissement des explications ou même de l'utilisation des citations des chercheurs hors de leur contexte. D'autres disent que parfois, ils ne reconnaissent même plus leurs articles après remaniements. Ils soulignent également qu'heureusement ces cas sont rares.

« Des fois je leur donne un sujet, c'est un peu mon bébé et en fait, ça ressort comme un autre enfant. C'est assez bizarre, c'est plus le mien en fait. Mais je leur en veux pas trop, ça va. je support...tolère la frustration. » P5

« Ils l'ont fait, mais on va dire qu'une fois que l'article a été changé à 60 %,

voilà. »P4

A l'opposé, d'autres ne s'en plaignent pas et ont conscience que leur texte peut et est souvent remanié, parfois même en mieux :

« Mes papiers sont donnés aux secrétaires de rédaction qui vont complètement les réécrire, les changer, les mettre dans un ordre différent, enlever des phrases, ajouter une nouvelle introduction, bref ils améliorent le papier même si parfois des erreurs se glissent. Mais en général, ils le rendent vraiment meilleur (...). Mais bon c'est le truc. Etre un journaliste ce n'est pas comme être un écrivain, ou un scénariste. On finit par devenir plus coulant avec notre prose quand on travaille tous les jours pour un journal et que notre texte est souvent modifié. Ca c'est l'autre problème quand on fait relire ces textes par des scientifiques. Il n'y a vraiment aucun intérêt pour moi à écrire 500 mots magnifiques, à les envoyer aux trois scientifiques que j'ai interviewés pour obtenir leur assentiment et accepter leurs modifications tout ça pour que quand je l'envoie pour publication ce soit changé en une brève de 20 mots en page 37. » P3

La question des remaniements a aussi mis en lumière les conflits qui peuvent exister entre les opinions des journalistes scientifiques sur la façon dont ils appréhendent ou couvrent un sujet et celles de leurs rédacteurs en chef. Ainsi une des journalistes se plaignait du poids que donnait son rédacteur en chef au principe de l'objectivité journalistique.

« (Je ne voulais pas citer quelqu'un) pro-écran et je veux pas qu'on induise le lecteur en erreur en disant que c'est bon pour les enfants alors que les recherches montrent que ça ne l'est pas. Il (rédacteur en chef) a dit ben si, par souci d'équité...et de justice, il faut qu'on présente les deux avis. » P5

Un autre indiquait que finalement ce qui comptait avant tout, pour pouvoir avoir son article accepté sans trop de remaniements, c'était de satisfaire leur rédacteur en chef. Des observations similaires avaient été faites dans la littérature (Cooper et al. 2002).

« Vous devez, votre motivation principale en tant que journaliste est de satisfaire votre rédacteur en chef, votre patron plutôt que de garder une vision plus large sur ce que vous faites» (en parlant de la qualité des résultats scientifiques couverts). S9

Dans l'ensemble, les journalistes que nous avons interviewés ne semblent pas voir leurs articles remaniés sur le fond mais plutôt sur la forme. Ils sont en général positifs dans leur appréciation de ces modifications et reconnaissent que la plupart du temps la qualité de l'article est améliorée. Même si dans certains cas, ces remaniements ont pu être source de conflits avec leur hiérarchie.

2.4.4- Choix du titre

Pour les journalistes que nous avons interviewés, comme pour la majorité des journalistes, les titres sont choisis par les rédacteurs en chef ou les secrétaires de rédaction.

« Les titres sont vraiment rarement écrits par les journalistes, en tous cas au Royaume-Uni. » S9

« En tant que salarié d'un journal je ne pense pas qu'on ait notre mot à dire dans le choix du titre. » P3

En général, les journalistes proposent un titre mais plutôt à titre indicatif. Le titre peut être retenu mais il est souvent changé. Pourtant, deux journalistes, un pigiste et un salarié, ont dit qu'ils choisissaient leur titre et que ceux-ci étaient la plupart du temps conservés. Mais, pour la plupart des journalistes, cela n'a en général pas d'importance. Ils savent que le titre et le chapeau sont décidés par les rédacteurs en chef et secrétaires de rédaction. Souvent ils pensent aussi qu'ils ne sont pas les mieux placés pour écrire leur titre et qu'une certaine distanciation face à l'article est nécessaire.

« Moi je suis pas très bon, j'aime pas y passer trop de temps, ça me saoule, quoi, tu vois, sur les titres...si t'as la bonne idée, enfin je mets toujours un titre tu vois, mais...y'a pas d'enjeu pour moi, y'a pas d'enjeu dans le titre réellement. » P2

« Ca c'est souvent la cuisine interne. On n'est pas très forts. On n'est pas forcément forts en titres. Et en titrairie en général. » P1

« Alors la titrairie c'est souvent quand même euh...des fois j'ai mon titre ils le prennent, des fois il est pas bon, j'ai pas d'inspiration, c'est la rédaction qui choisi. C'est variable. » P7

Cependant, certains journalistes ne sont pas toujours satisfaits du titre qui accompagne leur article et déplorent parfois le sensationnalisme qui y est associé :

« C'est assez peu négociable et ça donne parfois des choses punchy, percutantes, et ça donne parfois des choses qui me font limite honte. J'ai un chef qui est punchy et percutant et qui est tellement sûr de lui, que parfois, il est délirant - mais c'est reconnu de tout le journal - et qui ne change pas d'avis. Ça a donné lieu à des conflits incessants et ma place... comme ma place n'est pas de titrer, eh ben, je ne peux qu'attendre le jour où j'aurais le droit de titrer moi-même, essayer d'être plus nuancée, voilà. » S1

« Ça aussi journalistiquement, on vous laissera pas trop écrire un titre vraiment très près, ce qu'on appelle mi-figue mi-raisin quoi. Alors ça, c'est l'abus des... l'excès du journalisme actuel, des titres courts, frappants, attrayants, attractifs, etc., et si possible tout noir ou tout blanc. Ça on adore ça. Et ça combien de fois, j'ai dit que la science elle est ni toute blanche ni toute noire. » S6

Nous avons demandé aux journalistes de manière plus explicite si ils avaient parfois été surpris ou choqués du titre qui avait été choisi pour leur article. Cette question a mis en lumière des conflits avec les rédacteurs en chef ou les secrétaires de rédaction. En effet, même si les journalistes sont mécontents, ils ont une marge de manœuvre très faible pour influencer sur le choix du titre.

« C'est quand même leur boulot. Mais ça m'est arrivé d'avoir de mauvaises surprises, c'est à dire des titres qui étaient, c'est pas le fait d'être racoleurs, c'est le fait d'être faux par rapport au texte. Qui laissaient entendre quelque chose qui finalement n'était pas... » P6

« Ouais, en général chez nous, les éditeurs ont beaucoup de pouvoir. Les services des éditeurs les SR (Secrétaires de Rédaction) hein, et les gens qui éditent le papier hein ont pas mal de pouvoir. » S13

« Si c'est un papier complexe ou sujet à controverse, en général je vais vérifier le titre et si le secrétaire de rédaction a écrit un titre que je n'aime pas, je vais essayer de le faire modifier. Tout le monde ne peut pas faire ça et ça dépend aussi du journal et de sa culture. Si vous essayez de faire ça au Sun, ça ne changerait pas grand chose. » S9

Les rédacteurs en chef interviewés au cours de l'enquête ont eux bien sûr une position différente. Leur but est d'attirer le lecteur et un titre accrocheur aide beaucoup.

« L'identité du journal est très marquée par les titres, donc, on les refait souvent. Les titres fournis par les journalistes donnent une indication, une orientation pour élaborer un nouveau titre, mais en général, on ne les garde pas. (...) Il faut que ce soit intéressant dès le départ. Donc, je remanie quasiment systématiquement toutes les attaques des papiers, et je donne du rythme et du dynamisme au papier, en général. » S12

« On n'essaie pas de rabaisser les papiers mais plutôt de les promouvoir en disant : bon ça c'est vraiment intéressant mais est ce que ça signifie aussi ? On essaie de développer et d'apporter cette dimension supplémentaire ces éléments de contexte. Et voilà c'est là que les titres entrent en jeu. En ce moment notre ligne éditoriale c'est d'écrire des titres amusants en les poussant au maximum sans avoir tort. On ne veut pas écrire des trucs faux. On ne veut pas écrire des choses qui ne sont pas vraies mais en même temps, on veut parler des choses les plus intéressantes, donc on joue des métaphores et autres. On ne veut pas exagérer ou déformer quoique ce soit. Mais en même temps, on veut se focaliser sur le truc le plus intéressant, et on veut le présenter de la manière la plus intéressante possible pour que les gens le lisent. Donc ouais, c'est un compromis. » S10

Au total, pour les journalistes scientifiques interviewés dans notre enquête, l'important c'est donc d'être fidèle à l'information et de coller le plus possible aux résultats de la recherche y compris dans le choix du titre, quitte à être vraiment factuel. Pour les rédacteurs en chef, ce qui compte, c'est d'attirer le lecteur avec des titres percutants au détriment de la conformité aux résultats scientifiques. Un des journalistes pigistes justifie ce problème d'exagération du titre en disant que toute exagération est rééquilibrée dans le texte de l'article lui-même et que le lecteur aura toutes les informations correctes à la lecture de l'article entier.

« Il y a aussi le problème du titre... dans la culture de la presse écrite il y a toujours eu, enfin le titre est écrit par quelqu'un d'autre, par le secrétaire de rédaction, parce que c'est le titre qui accroche le lecteur, mais si vous lisez le papier jusqu'au bout, vous aurez la vision globale dans son entier, et l'exagération du titre sera un peu contrebalancée parce qu'il y aura d'autres points de vue, d'autres informations. » P3

Cette justification est cependant discutable puisque la majorité des lecteurs s'arrête au titre et au premier paragraphe, ce qui est d'ailleurs la raison d'être de l'écriture en pyramide inversée. Donc si le lecteur s'arrête au début de l'article, il n'aura pas les informations pour contrebalancer la survente des résultats dans le titre. L'idée qu'il se fait des avancées de la recherche est donc biaisée.

2.5- Interactions avec les scientifiques

2.5.1- interview des auteurs des études scientifiques

Tous les journalistes interviewés contactent au moins un des chercheurs qui ont réalisé l'étude pour écrire leur papier (voir paragraphe 2.4.2). Si la lecture de la publication scientifique n'est pas indispensable pour tous, les interviews des chercheurs impliqués dans l'étude sont elles primordiales pour apporter un contexte à l'étude qu'ils couvrent, pour mettre en lumière certains éléments ou comme le disait un pigiste, *« pour lui donner un léger vernis scientifique »*.

Pour certains des journalistes français, cette importance de pouvoir interviewer un chercheur a eu une incidence indirecte sur le choix de leurs sujets. En effet, leur maîtrise de l'anglais n'étant pas suffisante pour pouvoir réaliser des interviews, ils choisissent de se tourner vers des spécialistes français.

« C'est que du coup, vu que j'ai un mauvais anglais, enfin que je maîtrise pas l'anglais comme il faudrait pour pouvoir faire des interviews en anglais, je me basais surtout sur des recherches françaises. Ça, c'est peut-être un biais, on va dire, du métier, c'est que, du coup, j'essayais toujours de trouver des sources qui me permettent de retomber sur au moins

un auteur français à qui je puisse poser des questions, même si c'est pas le premier auteur, mais voilà, qui puisse m'informer sur ça. » P4

« Non, c'est compliqué, moi j'ai besoin de... j'ai besoin de parler en français à quelqu'un pour avoir des explications de base et je sais quand même qu'il y a un moment où ça va être plus facile en français, donc s'il y a un auteur français, je peux appeler l'auteur français. » S3

Les journalistes développent donc et entretiennent des relations de confiance avec les scientifiques qu'ils ont interviewés. Ils leur permettent de trier aussi dans le flot d'informations qu'ils reçoivent et d'éviter des écueils.

« Parfois, il y a des très mauvais papiers qui sortent, qui sont assez séduisants, qui « journalistiquement » ont un intérêt. Enfin, on pense qu'ils ont un intérêt. Et on le fait lire à quelqu'un de la communauté en qui on a plus ou moins confiance. Et puis, cette personne peut nous dire aussi : Ben, c'est nul. » S13

Ces observations sont en accord avec la littérature : les journalistes scientifiques entretiennent des relations privilégiées avec les scientifiques. Ils contactent les chercheurs qu'ils connaissent ou auxquels ils font confiance pour leur demander leur opinion sur la validité et la qualité d'une publication qu'ils ont l'intention de couvrir.

2.5.2- Rapport avec les chercheurs

Au cours des entretiens avec les journalistes, il est apparu que les relations entre les chercheurs et les journalistes avaient évolué et que dans l'ensemble, les journalistes étaient plutôt contents des interactions qu'ils avaient eues avec les scientifiques. Ainsi ces deux journalistes salariés expliquent qu'ils se sentent désormais plus à l'aise pour interviewer les chercheurs et qu'ils ont constaté une évolution positive de la disponibilité de ceux-ci depuis le début de leur carrière.

« Je n'ai plus la trouille d'appeler un chercheur, en fait. Je n'ai plus la trouille de me sentir méprisée. Bon après, c'est aussi que j'ai plus d'expérience donc peut-être, c'est une interaction, enfin, je ne sais pas comment ça se... la plupart des chercheurs que je rencontre, je dirais qu'il y en a 10 % qui sont obscurs et incapables de vulgariser et 90 % qui sont super-sympa, qui sont généreux de leur temps et qui vulgarisent bien. » S1

« Les scientifiques sont plus ce qu'ils étaient, c'est-à-dire qu'ils communiquent beaucoup mieux qu'avant, beaucoup mieux. L'image du scientifique qui a des toiles d'araignées sous le bras, ça, c'est terminé. Ils sont vraiment... La plupart du temps, quand je les ai au téléphone, je suis toujours très surpris de l'aisance de la... avec laquelle ils s'expriment parce que ce sont des gens qui... » S8

D'autres journalistes indiquent qu'ils apprécient réellement leurs interactions avec les scientifiques et que ceux-ci sont enthousiastes à l'idée de vulgariser leurs travaux.

« Enfin, je n'ai pas eu l'impression qu'il y ait des scientifiques qui essayaient de rendre leur truc fantastique, s'il ne l'était pas ou... non, je les trouve assez honnêtes et assez... en général, assez agréables et capables de vulgariser, d'expliquer et... oui, moi, j'ai toujours trouvé que les scientifiques français avaient une grande ouverture d'esprit, enfin, j'étais toujours étonnée par leur... même leur accessibilité. » S4

« Globalement, moi je trouve que les scientifiques sont toujours ravis de parler de leurs travaux et d'expliquer ce qu'ils font et qui, en général, certains ont peut-être plus de difficultés que d'autres à... à utiliser des modes simples. Mais globalement, peut-être qu'ils font tous l'effort, de toute façon, et qu'ils ont envie de partager. » S3

Certains ont quand même une position plus critique, notamment vis à vis des médecins. Ils pensent que certains chercheurs les utilisent pour promouvoir et exagérer les implications de leurs travaux.

« J'ai l'impression qu'il y a beaucoup de scientifiques qui veulent, comment dire, enfin, d'études sensationnalistes un peu... des scientifiques qui veulent faire le buzz. Alors, je ne sais pas, c'est une impression que j'ai, mais... on veut se faire médiatiser, etc., et je trouve ça vraiment dommage, parce que quelque part, derrière, enfin, le scientifique en question perd de la crédibilité. » S5

« Mais énormément, c'est... c'est affreux, mais je me méfie énormément des médecins, parce que, rares sont ceux, justement, qui ont une veille scientifique ; et, sur certains sujets, enfin, je les interviewe quand il faut interviewer un médecin, quand le sujet est d'interviewer un médecin ou une technique médicale. Mais, je passe toujours derrière pour voir si la technique qu'ils mettent en avant, comme étant celle qui est en pointe, celle qui a les meilleurs résultats a, effectivement, les meilleurs résultats d'un point de vue de la littérature médicale. » S7

Cette relation particulière que les journalistes scientifiques entretiennent avec les scientifiques modifie également leur position vis à vis de la relecture de leur texte. En général, les journalistes ne font pas relire leur texte ni les citations, c'est le principe de la liberté et de l'indépendance de la presse. Mais les journalistes scientifiques interviewés dans cette enquête ont une position différente comme le montrent ces citations de deux journalistes salariés et d'un pigiste.

« Je ne faisais jamais valider les propos (...) j'étais absolument farouchement contre, avec des arguments et j'étais intraitable (...). (Puis cette journaliste a modifié ses positions du fait de ses rapports avec les chercheurs) Je me dis : « Tiens, dans leur discipline, j'ai

l'impression que les articles scientifiques, ils naviguent, ils bougent, ils se font relire. » Je me suis dit « s'ils n'avaient pas tort, et si dans cette page-là, qui est sanctuarisée, qui n'est pas : « je critique la recherche, mais j'essaie de raconter ce que fait la recherche, là, je vous disais que je n'ai pas tellement de... il n'y a pas tellement de... il n'y a pas de conflit possible », donc, j'ai décidé que je le faisais [relire]. Au début, j'envoyais que les citations en me disant : « Ils ne vont pas quand même tout relire. » Et puis, en fait, je pense que ça m'évite d'écrire des bêtises sur des... donc, je le fais. (...)Ça n'a pas été simple, je me suis faite engueuler plusieurs fois par mes chefs qui m'ont dit : « Je ne comprends pas pourquoi tu fais relire tes trucs, l'indépendance, c'est sacré. » S1

«Si ils me le demandent, je suis toujours d'accord pour leur [faire] relire (...). Mais il n'y a pas de présomption, il n'y a jamais eu de présomption que quelqu'un qui donne une interview à un journaliste pourra donner son aval sur ce qui est écrit, en tous cas au Royaume-Uni. C'est vraiment à l'opposé de l'idéal des journaux. Si vous écrivez à propos de quelque chose, si vous parlez à quelqu'un, ils n'ont absolument pas la possibilité de valider ce que vous écrivez parce qu'on ne voudrait pas voir ça avec des hommes politiques, des footballeurs ou des hommes d'affaires, ni pour des gens impliqués dans des procès. La science c'est différent, parce que, en général, un entretien entre un journaliste et un scientifique c'est une exploration, c'est une ouverture d'esprit ce n'est pas une confrontation. » P3

« Oui, c'est très spécifique au journalisme scientifique, ça. Moi, j'ai été formée à Science et vie junior, j'ai travaillé dans plein de publications scientifiques, à Science et vie... tout ça, et j'ai toujours, oui... j'ai toujours fait valider mes... mais, je vais même vous dire plus loin, non seulement je fais valider les citations. Si c'est vraiment un papier polémique, je ne leur fais pas lire l'intégralité du papier, mais la plupart du temps, je fais lire l'intégralité du papier. » S12

Seuls deux journalistes ont indiqué qu'ils ne laissaient pas les chercheurs relire leur texte, ne serait-ce même que la citation. Le premier est un journaliste et éditeur britannique très attaché à son indépendance et qui tient à ce que « l'histoire » soit celle du journal et non celle du scientifique. Le deuxième travaille pour la télévision et les émissions ne sont jamais montrées avant diffusion.

« ça c'est un principe journalistique, parce qu'on n'est pas une entreprise de communication, c'est-à-dire qu'on ne fait pas la com, ni la publicité de personne, donc, on prend nos responsabilités pour dire les choses telles qu'elles sont. Et du coup, on n'a pas besoin de les faire valider par la personne qui les a dites. (...) On ne montre jamais le document avant la diffusion. » S11

Neuf journalistes laissent relire les citations mais uniquement à la demande des scientifiques. Les dix autres laissent les scientifiques relire les citations et leur texte, notamment parce que cette vérification leur permet d'éviter des erreurs.

« Et puis, en fait, je pense que ça m'évite d'écrire des bêtises sur des... donc, je le fais. » S1

« S'il est un peu polémique, controversé et que je met plusieurs avis en même temps, que je les confronte, là, euh, bien évidemment non, je fais pas relire, mais si c'est pas le cas, je fais relire. Au moins pour qu'il me donne son avis et pour pas que j'aie fait de contresens dans le reste de l'article. » P6

La plupart des journalistes sont assez favorables à cette relecture, au moins des citations voire du texte entier parce qu'ils comprennent la position des chercheurs qui veulent se rendre compte de l'utilisation qui est faite de leurs propos

« Parce que je me dis que c'est déstabilisant pour elle si en plus c'est des fois des choses très complexes donc c'est déstabilisant pour elle si y'a quelque chose qui paraît en son nom qui n'est pas juste, en clair qui trahit sa pensée, donc là je fais vraiment attention de toujours faire relire les propos à la personne. » P5

« Ben moi ça m'est arrivé plusieurs fois de pas relire et de me dire « mais c'est absolument pas ce que j'ai dit ». Donc je me dis, cette violation de la...cette transformation de la parole, elle est super chiante, je comprends très bien qu'un expert qui voit son nom associé à des propos qu'il a pas tenus ou qu'il a pas tenus comme ça, moi ça choque absolument pas. Enfin pour moi c'est pas du tout un problème de liberté de la presse, de pensée, de machin. » S2

« Ça ne me choque pas de faire relire les citations, à partir du moment où moi j'ouvre les guillemets, c'est censé être vraiment ce qu'ils disent, donc là ça ne me choque pas et en général, ça se passe bien. D'ailleurs, je n'ai pas de problème avec ça. » S3

Parmi les journalistes qui laissent relire les citations, certains ne le font pas volontiers : ils pensent que les chercheurs voudront alors modifier leur propos et que ce n'est pas cela qu'ils attendent des chercheurs lorsqu'ils les interrogent et leur demande leur opinion.

« Des fois il y en a qui modifient complètement ce qu'ils ont dit, ça c'est très agaçant. (...) Si je prends des notes de ce que vous me dites, un moment sur un sujet où vous êtes assez passionnée etc. vous allez dire des choses, vous allez monter d'un tout petit degré supplémentaire. Je vais vous le faire relire, je suis presque sûr que les différentes raisons, vous... voilà une jeune personne en début de carrière, ou un... Vous allez un petit peu abaisser le niveau de... de virulence ou d'emportement ou de dynamisme dans vos propos.

Donc, le malheur c'est que souvent, ils atténuent ce qu'ils ont dit. Donc, ça c'est un peu embêtant, et ça ça me gêne parce que s'ils l'ont dit, c'est qu'ils le pensent quand même. » S6

« Mais c'est très bien de se laisser emporter ! C'est ça qu'on veut, parce que ça montre qui ils sont et ce qu'ils pensent ? Et si vous leur faites relire leurs citations et bien ce qui arrivent c'est qu'ils vont vouloir les réécrire et ils le font d'une telle manière que le papier n'est plus lisible, il est ennuyeux technique, on dirait un article scientifique, ils ont enlevé tout ce qui était intéressant parce que les scientifiques sont avant tout des scientifiques. » S10

On voit donc que les rapports entre les scientifiques et les journalistes sont un peu déséquilibrés : les journalistes accordent un grand crédit aux scientifiques, leur laissent relire leur texte pour leur éviter des erreurs, des contresens, autant de pratiques que les journalistes généralistes condamneraient au nom de l'indépendance de la presse - comme le font d'ailleurs remarquer plusieurs journalistes. Ils conviennent également que cette différence est particulière au journalisme scientifique. Il est vraisemblable que cette attitude soit exacerbée par la dépendance des journalistes scientifiques à leurs sources. Ils essaient donc de ne pas les contrarier. Comme le disait un des journalistes pigistes (P3) *« le boulot d'un journaliste scientifique c'est de refléter ce que le chercheur pense et dit, ce n'est vraiment pas de le pousser à tenir des propos exagérés »*. Cette déférence vis à vis des chercheurs est sans doute aussi indirectement impliquée dans la qualité médiocre de l'information scientifique présentée dans les médias.

2.6- Opinion sur la reproductibilité des sciences

Dans le champ scientifique, le problème de la reproductibilité des études biomédicales est connu et a suscité de nombreux éditoriaux, recherches et propositions dans les publications scientifiques (Begley et al. 2015; Ioannidis 2005, 2005b, 2008; Macleod et al. 2014). Il l'est moins dans le champ journalistique où les articles publiés sur le sujet sont moins nombreux et récents (Lehrer 2010; The Economist 2013b). Un journaliste vient de publier un livre sur ces problèmes de reproductibilité en recherche biomédicale (Harris 2017). Ces articles et cet ouvrage laissent penser que cette crise de la reproductibilité et ses causes commencent à être connues des journalistes. Une des questions de l'enquête était justement de connaître la position des journalistes scientifiques sur cette question de la reproductibilité. La question était la suivante *« A votre avis, dans le domaine de la recherche biomédicale, quelle est la proportion de découvertes initiales qui ont été soit réfutées soit fortement atténuées par les études ultérieures sur la même question? »*. Nous voulions savoir si les journalistes interviewés avaient conscience que la majorité des études initiales sur un sujet étaient réfutées

par les études ultérieures. Les réponses à cette question sont présentées sur la figure 4.

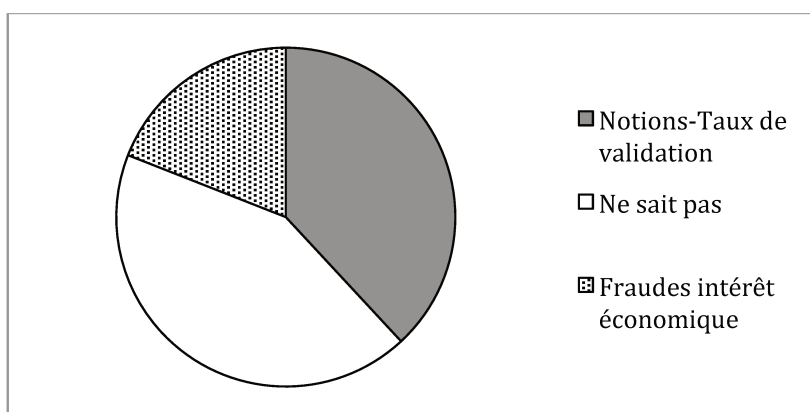


Figure 4 : Réponses des journalistes quant à leur opinion à propos de la validité des résultats de la recherche biomédicale

Huit des journalistes interviewés ont une opinion sur la reproductibilité des résultats de la recherche biomédicale. Deux d’entre eux, des journalistes salariés, parlent des études qui se contredisent et donnent notamment l’exemple du sucre qui, en fonction des études, est bon ou mauvais pour la santé. Les six autres (quatre salariés et deux pigistes) connaissent le problème et ont mentionné des taux de reproductibilité des résultats de la recherche entre 40 et 70%.

Neuf des journalistes interviewés ont répondu qu’ils ne savaient pas. Après discussion avec eux et après leur avoir donné quelques exemples, quatre journalistes (trois pigistes et un salarié) ont mentionné des taux de reproductibilité entre 50 et 90%. D’autres (quatre salariés) n’avaient vraiment pas d’opinion et ont alors parlé des fraudes et des « *des débats scientifiques qui n’en étaient pas vraiment mais étaient plutôt des débats politiques* » (exemples donnés : industrie du tabac et réchauffement climatique). Enfin un des journalistes (salarié) a dit qu’il ne voulait pas se prononcer.

« La question est très difficile à répondre, c’est trop difficile ça. Parce que dans combien de cas la découverte est reproduite ? J’essaye de pas trop me prononcer, enfin de pas... je n’ai pas à me prononcer là-dessus, enfin même si on y croit, on y croit pas un peu. »

J6

Cette question a été assez problématique pour quatre journalistes. Deux ont associé reproductibilité avec rétractations et fraudes. Ils ont donné des exemples de fraudes, comme l’affaire des cellules souches de Hwang, l’étude de Séralini sur les OGM ou des cas de pseudosciences comme le clonage de Raël. Ils ont aussi mentionné l’importance du site répertoriant les rétractations d’articles scientifiques (retractionwatch.com). Un troisième a d’abord pensé aux essais cliniques, à leur faible taux de réussite, aux intérêts économiques et aux conflits d’intérêts qui entrent en jeu. Le dernier a parlé du temps nécessaire pour qu’un

sujet prometteur devienne un outil.

« Alors ça je dirais que c'est le plus difficile à vérifier. Y'a que l'épreuve du temps, en quelques sortes, qui...même si le temps se raccourcit avec les nouveaux moyens de communication...mais on peut le voir parce que...même si c'est pas une garantie, mais on peut le voir par exemple...l'un des principes de la recherche scientifique, c'est que le travail doit être reproductible. » P6

Au total, 12 des 21 journalistes interviewés ont une certaine conscience des problèmes de reproductibilité des résultats de la recherche biomédicale. Même si l'un d'entre eux était vraiment trop optimiste et pensait que seuls 10% des résultats n'étaient pas reproduits, les autres journalistes ont proposés des taux de validité en accord avec nos observations (le taux de reproductibilité varie entre 10 et 70 % en fonction des domaines étudiés). Curieusement, le fait d'avoir une formation scientifique ne semble pas influencer la connaissance, ni faciliter la prise en compte de ce problème. En effet, seulement deux des dix journalistes diplômés en sciences ont répondu directement à la question en donnant des taux proches de la réalité moyenne. Cinq ont dit qu'ils ne connaissaient pas la réponse et trois ont mentionné les fraudes et les rétractations en réponse à la question. Près d'un tiers des journalistes scientifiques confondent donc les problèmes de reproductibilité avec les fraudes. Ils n'ont pas conscience du fonctionnement normal de la recherche, qui est un processus cumulatif évoluant d'une observation initiale incertaine vers un consensus validé par la reproduction des résultats par plusieurs laboratoires indépendants.

En revanche, les journalistes sont unanimes sur la nécessité d'informer le public en cas d'invalidation d'une étude qui aurait été médiatisée.

« Oui, mille fois oui. Mais même à tout niveau, c'est-à-dire que même si elle n'a pas été réfutée, mais juste remise en contexte, si je peux faire un article dessus, oui. » S7

« Oui, bien sûr, oui. Oui, mais d'ailleurs, c'est comme ça que c'est traité par la presse en général. Enfin, lorsqu'il y a une honnêteté intellectuelle sur... C'est comme ça que c'est traité par la presse en général. » S8

« C'est indispensable. C'est indispensable de faire le suivi, surtout des gros sujets qui ont été très médiatisés, c'est indispensable de faire le suivi de toutes les étapes de rétractation, de doute, d'enquêtes, tout ça, il faut le faire. » S2

Ils sont aussi assez critiques sur ce non-suivi des études, pour eux c'est quelque chose qui devrait être fait systématiquement.

« Et ça, c'est un truc qui m'a toujours un peu étonnée, c'est qu'on écrit un article, une brève ou quoi que ce soit, mais on effectue pas de suivi sur ce sujet-là et on passe tout de suite

à autre chose. Et donc, du coup, je serais bien incapable de vous dire quelles ont été les suites de tous les articles que j'ai faits en première... Voilà. » P4

« Mais en tout cas, je pense que c'est un... c'est souvent un problème pour le journalisme scientifique, c'est d'avoir ce suivi-là justement. C'est d'être capable de ne pas être naïf sur un sujet et de ne pas justement monter en épingle quelque chose qui n'est pas suffisamment costaud ou qui n'est pas vraiment nouveau. » S3

Ils concèdent néanmoins que leurs rédacteurs ne leur donneraient certainement pas la place nécessaire pour écrire sur le sujet.

« J'aurais du mal à leur donner autant de place. Je sais que... je n'ai pas d'exemple concret en fait, mais je pense que je n'obtiendrais pas autant de place pour les réfuter. Mais j'ai une demande d'exclusifs, de scoops médicaux, de la part de mes chefs. Et j'essaie, à chaque fois, de leur répondre que la recherche, elle progresse souvent à petits pas, qu'en particulier - je pense au domaine de l'Alzheimer- que c'est nuancé, qu'il y a des désaccords dans la recherche donc, voilà, j'essaie de faire ce genre de journalisme médical prudent, nuancé, de donner la parole au contradictoire et je vois bien que mes chefs aimeraient mieux que je dise qu'on va guérir les gens demain Sans jamais dire après-demain qu'on s'était planté et qu'on n'allait pas les guérir de sitôt. » S1

D'autres journalistes ont dit qu'ils n'avaient pas ou peu été confrontés à ce problème, qu'ils n'avaient jamais eu à écrire un papier sur un sujet qu'ils avaient couvert et qui avait été invalidé par la suite. Ils l'expliquent en disant qu'ils se méfient des résultats annoncés à grands fracas et qu'ils attendent de voir plusieurs études sur un même sujet, un « sillon » pour écrire sur le sujet.

« Je trouve que pour informer le grand public - puisque moi, c'est ce que j'ai fait en gros, quoi -, il vaut mieux attendre que ça soit confirmé, au moins par deux ou trois études, quoi ! Mais bon après, il m'est peut-être arrivé de faire une étude sur un sujet, c'était un peu ténu, mais en général, je faisais attention à ça. Ou aussi, que par exemple, ce soit testé sur une grosse population ; que l'étude, elle soit sur une grosse population, que ce soit pas genre sur 300 pelos. Quand il y a toutes les études qui sont sur plus de 2 000 personnes, voilà, ça devient plus... Donc, je faisais attention à ces éléments-là, oui, effectivement. » S4

« En général, j'essaie de mettre les bémols quand ils s'imposent, mais moi, je suis plutôt conservateur et je fais très peu de sujets sur des totales nouveautés. C'est-à-dire les trucs qui vient de sortir, qui casse la baraque et très étonnant. En général, j'essaie de pas trop traiter ces trucs-là. » S13

Enfin plusieurs journalistes se sont un peu dédouanés de ce problème de non-suivi des invalidations des études couvertes en critiquant le fonctionnement des médias. Quelques citations illustrent bien cette position :

« Là, vous postuleriez que le doute peut avoir sa place dans les journaux et je pense que le doute, il a sa place dans les classes de philosophie, j'espère, mais dans les journaux peu. D'ailleurs, si ces journaux ne vont pas bien, c'est peut-être en partie à cause de ça, donc c'est des choses que j'entends. Si un jour, j'étais chef de service santé-médecine, je pense que vous me donnez l'idée, mais ça serait intéressant, pas de donner... je ne vais pas non plus me transformer en conférence critique, mais en tout cas, je donnerais moins de place, mais je donnerais une petite place, en fait. Ça ferait sûrement vivre le débat et j'imagine que le lecteur, il aurait plus confiance en nous, si on acceptait de reconnaître qu'on se plante parfois. » S1

« Finalement le but de la presse, des sites web c'est de fournir une quantité d'informations que le public va avoir envie de lire. Et c'est bien dommage mais les papiers sur des résultats solides, ou ceux qui confirment ce qu'on sait déjà ou alors des papiers sur le fait que quelque chose qu'on pensait extraordinaire n'est en fait rien du tout, ces papiers là dans le fonctionnement actuel des médias auront moins de chance d'être couverts que les études initiales (...). Et puis il y a, je pense, une méconnaissance totale du fonctionnement de la recherche dans la société, mais bon c'est la nature humaine, malheureusement, ce qu'on veut ce sont des certitudes. » P3

« Mais bon, si on pense aux incitations pour les journalistes ? Finalement qu'est ce que ça fait si ils présentent des trucs faux ? Si ils n'agissent pas comme vous le suggérez (prudents dans la présentation des résultats de la recherche, réalisant un suivi des études qu'ils ont couvertes) ? Ou si au contraire certains le font ? On a besoin d'arriver à un point où plusieurs personnes se plaignent, où les lecteurs refuseront d'acheter, enfin je veux dire, de s'intéresser aux études initiales qui ne sont pas fiables. » S9

Les journalistes, même s'ils déplorent le sensationnalisme et les exagérations qui accompagnent les articles sur les découvertes scientifiques, reconnaissent qu'ils ne peuvent pas faire grand chose pour modifier cet état de fait. Ils mettent en cause l'organisation de l'activité journalistique - tournée vers la quantité et la rentabilité plutôt que vers la qualité et la fiabilité - et l'absence de répercussions si l'information présentée est erronée. Autrement dit, ils conviennent qu'il serait important de tenir compte de l'incertitude associée aux résultats initiaux qu'ils couvrent mais en pratique disent que cette incertitude n'a pas sa place dans les articles de presse.

A l'opposé d'autres ne s'inquiètent pas de présenter des résultats d'études initiales comme plus certains qu'ils ne le sont puisque leur rôle n'est pas d'être des « vitrines pour la science ».

« C'est qu'on n'est pas là pour expliquer la science. On est là pour expliquer le monde, décoder le monde et on va aller chercher l'information scientifique pour voir comment cette avancée des connaissances... Parce que moi, la science, pour moi, c'est ça, sa définition, c'est l'avancée des connaissances. » S14

« Ce n'est pas notre rôle de promouvoir la science et de s'assurer que le grand public a une bonne opinion de la science. » S10

« Moi, je crois que écrire, c'est critiquer le pouvoir, de manière générale quoi. Et donc moi, je pense que mon rôle, c'est de parler de la science quand elle interroge les pouvoirs en place. Parce que c'est ça qui fait avancer le processus démocratique. » S13

La moitié des journalistes semble être familière avec les problèmes de reproductibilité des études initiales. La majorité convient qu'il est important d'informer le public quand des études scientifiques couvertes par la presse ont été invalidées par la suite. Il ressort toutefois qu'ils ne semblent pas tenir compte de l'incertitude lorsqu'ils couvrent des études initiales. En tous cas, leurs opinions vis à vis de l'incertitude scientifiques sont variables. En effet, certains déplorent le manque de « suivi » de la presse, le manque de temps et d'espace pour traiter correctement des sujets. D'autres, au contraire, pensent que le rôle des journalistes n'est pas de « promouvoir les sciences » mais de les présenter de manière critique – même si on peut se demander où est leur esprit critique lorsqu'ils recopient les communiqués de presse – et de satisfaire les lecteurs et non la communauté scientifique.

2.7- Réflexions sur leur rôle

La question de la reproductibilité des recherches biomédicales se situait dans la dernière partie de l'interview mais avant une question sur le discours des scientifiques : s'adressent-ils de la même manière à leurs pairs et aux journalistes ? Cette configuration a amené certains journalistes à réfléchir sur leur rôle et leurs pratiques. Ils ont insisté sur le fait que leur rôle était celui de pourvoyeur d'informations mais aussi qu'ils n'étaient pas les seuls responsables dans les cas où l'information est biaisée. Selon certains, les institutions de recherche sont autant à blâmer que les médias.

« Mais c'est aussi pour ça que je fais ce boulot, c'est de donner le maximum de connaissances sur ce sujet. Et c'est je pense mon rôle essentiel. Après je ne peux pas jouer au justicier. C'est pas... c'est pas mon rôle non plus puisque finalement, y'a quand même des institutions qui doivent ou devraient le faire. » P6

« Ce qui fait que ça a donné, finalement, le résultat des courses, à cause d'un manque d'investissement peut-être en amont, un manque d'information sur...qui viennent du net, même, des éléments de revues sérieuses, euh, ben le public est dupé en fait, il a une mauvaise information. Donc il est désinformé. C'est vraiment... » P5

D'autres ont eu des réflexions plus larges sur le fonctionnement de la presse, sur le fait que souvent il n'y ait pas de remise en question ni de remise en contexte des propos tenus dans les médias.

« Le journaliste, il est sûr de lui, il a raison. C'est le petit gosse de CM2 qui lève le doigt devant la maîtresse et qui est sûr qu'il a tout compris, en fait. C'est son défaut professionnel. Les chercheurs en ont peut-être d'autres, mais le défaut de ma corporation, c'est qu'elle se remet pas en cause et qu'elle veut être la première partout. Elle veut du scoop, elle veut du neuf et elle n'a pas de mémoire. » S1

« (A propos d'un tabloïd anglais) Donc ils couvrent déjà les sciences mais personne ne connaissait rien à rien ; c'était un truc aliène pour eux. Donc voilà, ils ne sont pas, c'est leur job et ils ne diront pas non aux sujets de science mauvais si ça peut plaire, si ils pensent qu'ils auront des clics, ils vont le publier non ? Donc voilà, je pense que c'est ça le problème. C'était plutôt la façon de fonctionner des médias, comment ils travaillent et qui ils emploient ». S10

Une pigiste a ajouté que la rémunération des journalistes pigistes ne permettait pas non plus de faire des investigations trop poussées.

« Non, enfin, vraiment, moi, je pense que le travail des journalistes, il est très lié à la façon dont on les paie, dont on les rémunère. C'est pour ça qu'on bosse avant tout! Et donc du coup, notre travail, il est directement lié à ce phénomène. » P4

Il semble donc qu'il existe une tension chez ces journalistes : ils se veulent critiques et indépendants, mais reconnaissent en même temps que l'évolution de leur profession - avec de moins en moins de journalistes scientifiques salariés et de plus en plus de sujets à traiter- les laisse vulnérables face à leur hiérarchie et à l'habileté des organisations scientifiques à mettre en avant les travaux de scientifiques en utilisant des critères d'intérêt médiatique.

III- Conclusion

Cette enquête a été réalisée afin d'explicitier les pratiques des journalistes scientifiques lorsqu'ils rapportent des résultats de la recherche biomédicale et de connaître leurs opinions quant à la reproductibilité de ces résultats. Dans un premier temps, nous revenons plus précisément sur ces éléments ainsi que sur ses limites de cette enquête. Nous mettons à l'épreuve la théorie du gatekeeping (Shoemaker et al. 1996), décrivons les différents facteurs

qui déterminent le choix des sujets et dans quelle mesure la sélection des sujets couverts dépend des journalistes scientifiques. Nous présentons ensuite la méconnaissance des journalistes de la façon dont les sciences sont produites. Enfin, nous questionnons le rôle de médias en tant que quatrième pouvoir.

3.1- L'enquête et ses limites

L'enquête auprès de 21 journalistes scientifiques a été réalisée dans le but de confronter les observations que nous avons faites lors de la quantification et de l'analyse de la couverture médiatique des résultats de la recherche biomédicale avec les pratiques des journalistes telles qu'ils les décrivent. En particulier, nous voulions essayer de comprendre (1) pourquoi la couverture des résultats de la recherche est si homogène et reprend les grandes lignes des communiqués de presse ou des agences de presse, (2) pourquoi les journalistes ne font pas état des invalidations des études qu'ils ont couvertes, (3) pourquoi, en relayant essentiellement des études initiales, ils présentent au grand public des résultats au mieux incertains, au pire faux et enfin (4) pourquoi ils présentent des résultats comme plus certains qu'ils ne le sont en omettant de mentionner la nécessité de confirmer les résultats et en exagérant les titres.

Dans cette enquête, nous avons observé deux types de journalisme celui des « actus » et celui des dossiers complexes. Les « actus » rapportent les résultats d'une étude scientifique précise et récemment publiée. Par contre, dans les dossiers, les journalistes s'efforcent de faire le point sur une question et font souvent appel à des experts aux opinions contradictoires. Les dossiers décrivent très rarement les résultats d'une étude précise et récemment publiée.

La décision des journalistes de traiter une information scientifique soit comme une « actus » soit comme un dossier peut en elle même poser problème. Ainsi, pour les journalistes de notre enquête, les OGM constituent évidemment une controverse méritant un dossier étoffé. L'un d'entre eux a rappelé les problèmes engendrés par une mauvaise vulgarisation des recherches sur les OGM:

« Voilà, la génétique, ça fait partie des sujets où il faut vraiment réfléchir à la manière de le vulgariser. Parce qu'autrement on peut se tromper et risquer énormément. L'histoire des OGM le montre bien. La peur par l'ignorance. » P6

Nous avons donc été très surpris que beaucoup de journalistes de notre enquête aient mentionné « Crispr/Cas9 » (Jinek et al. 2012) en parlant d'un outil révolutionnaire et qu'aucun n'ait parlé des implications éthiques potentielles qui entourent l'utilisation de « ces ciseaux génétiques ». Pour eux cette nouvelle technique semblait donc une « actus » sans

problème alors même que l'une des auteurs de cette découverte a publiquement fait part de ses inquiétudes quant aux applications.

Ces observations rappellent les deux types de journalismes que Marchetti avait décrits pour le champ médical: le « journalisme de routine » qui correspond somme toute ici aux « actus » où le journalisme suit et fait confiance aux institutions scientifiques, et le journalisme « de rupture » qui s'inscrit en contre-pouvoir et qui serait ici représenté par les dossiers complexes pour lesquels le journaliste enquête et critique.

Cependant, les 426 articles de presse dont nous avons analysé le contenu peuvent tous être considérés comme des « actus », même si quelque uns allaient au-delà de la simple description d'une étude en évoquant par exemple les questions éthiques qu'elle soulève. Nous n'avons pas pu analyser le contenu d'articles de fond ou de dossiers complexes. Ceci tient à notre méthode de recherche des articles de presse puisque nous avons collecté les articles couvrant une étude précise, définie par avance et venant juste d'être publiée. Notre analyse de contenu ignore donc complètement la part de travail que les journalistes considèrent pourtant comme la plus noble: la rédaction de dossiers complexes.

Une autre limite de cette enquête réside dans la population des journalistes interviewés. En effet, il aurait été intéressant de pouvoir interroger les journalistes scientifiques qui étaient les auteurs des articles de presse dont nous avons analysé le contenu (chapitre III). Cet objectif était en pratique très difficile à réaliser puisque ces articles de presse ont été publiés entre 1988 et 2009, qu'ils n'étaient pas tous signés et que ces articles étaient issus de la presse anglo-saxonne (Royaume-Uni, Etats-Unis, Canada, Australie et Nouvelle-Zélande). Certains de ces journalistes travaillant au Royaume-Uni ont cependant été contactés mais aucun n'a donné suite à nos demandes. Dans la discussion ci-dessous, nous faisons tout de même implicitement l'hypothèse que les commentaires des journalistes interviewés pourraient éclairer des pratiques communes à tous les journalistes scientifiques. Ces commentaires pourraient donc expliquer certaines de nos observations relatives au contenu des articles analysés au chapitre III. Cependant, la validité de ces explications est limitée par la solidité toute relative de notre hypothèse. Cette limitation doit d'autant plus être prise en compte que notre enquête a été réalisée auprès d'un petit nombre de journalistes scientifiques, principalement français, et dont la sélection a été largement arbitraire.

3.2- Influences des différents niveaux du *Gatekeeping*

Les questions de notre enquête, tournées vers les pratiques et routines des journalistes scientifiques, ne permettent pas d'analyser l'influence des facteurs externes et sociétaux. Par contre, nous avons pu étudier les trois niveaux d'influence internes aux médias : les

caractéristiques individuelles des journalistes, les routines de communication et les caractéristiques organisationnelles propres aux institutions médiatiques.

3.2.1- Facteurs individuels

Cette question n'a pas été abordée directement dans l'enquête. Pour les journalistes français cependant, leurs connaissances, notamment leur maîtrise plus ou moins grande de l'anglais, a clairement limité leurs choix. En revanche, aucun des journalistes n'a vu comme un problème de ne pas avoir de formation en sciences pour traiter un sujet. En effet ceux qui n'avaient pas de formation en sciences travaillaient plus leur sujet en amont avant d'aller interviewer des chercheurs qui, au besoin, pouvaient leur expliquer des points obscurs, et ceux qui avaient une formation scientifique universitaire prétendaient ne pas en avoir lorsqu'ils interviewaient des scientifiques afin d'obtenir des réponses compréhensibles par leurs lecteurs. Les journalistes ont cependant indiqué que certains sujets étaient plus compliqués et techniques que d'autres, mais cette constatation était faite autant par des journalistes avec une formation scientifique que sans.

Dans leur étude sur les journalistes scientifiques de la presse allemande, Guenther et Ruhrmann ont relevé deux facteurs individuels qui influençaient le choix des sujets : l'intérêt personnel des journalistes pour le sujet en question et la façon dont les journalistes interprétaient leur rôle (Guenther et al. 2013). Or, dans notre propre enquête il n'est pas apparu clairement que les journalistes suivaient et donc couvraient des sujets plus que d'autres. Certains étaient plus tournés vers les sciences humaines, d'autres vers l'environnement mais ils couvraient cependant tout un panel de sujets différents. La familiarité avec le sujet était donc un critère moins important que sa nouveauté et sa facilité de traitement.

Les journalistes interviewés dans cette enquête percevaient leur rôle comme celui de pourvoyeur d'information, mais insistaient également sur leur rôle d'enquêteur qui se doit « *d'avoir ce recul et de mettre des bémols (...). De mettre un regard critique, même s'il est persuadé que le sujet paraît intéressant* » (P6). Ceci ne reflète pourtant pas ce que nous avons observé dans l'analyse de contenu des articles de presse (chapitre III). La couverture médiatique d'une même étude scientifique était relativement homogène ; on retrouve des phrases identiques et les mêmes citations des mêmes chercheurs d'un article à l'autre. Ce regard critique, ces bémols sont en réalité peu présents dans les articles que nous avons analysés.

Les données de notre enquête, qui n'interrogeaient pas directement les journalistes sur ce qui les poussaient à choisir un sujet plutôt qu'un autre, suggèrent que les caractéristiques

individuelles des journalistes, telles que la formation ou le cadre professionnel, n'entrent pas beaucoup en compte dans le choix des sujets. Pour les journalistes interviewés, ce sont plus les critères de nouveauté et de simplicité qui semblent importants.

3.2.2- *Le poids des routines*

Notre enquête montre également que les journalistes s'appuient sur des sources qu'ils ont l'habitude d'utiliser et sur la confiance qu'ils accordent à la fiabilité des résultats présentés dans les journaux scientifiques prestigieux pour choisir leur sujets. Cette observation est en accord avec les résultats décrits au chapitre précédent (chapitre III) ainsi que ceux de la littérature : les journalistes favorisent les études publiées dans des revues scientifiques prestigieuses et accompagnées de communiqué de presse. A l'inverse, des études scientifiques publiées dans des revues à comité de lecture moins prestigieuses et n'ayant pas fait l'objet d'un communiqué de presse n'ont pratiquement aucune chance d'être couvertes. Cette facilité d'utilisation explique ce que nous observons dans la couverture médiatique des études scientifiques de notre base de données : elle concerne des sujets peu controversés et sont publiées dans des journaux scientifiques prestigieux. Les journalistes font alors confiance aux résultats de ces études et aux chercheurs. Même lorsqu'une même question scientifique est couverte plusieurs fois, les journalistes ne donnent peu ou pas d'éléments de contexte et se contentent de parler « d'autres études contradictoires ». Par contre, ils présentent les résultats de la nouvelle étude qu'ils couvrent comme indiscutables et suffisants pour apporter une conclusion définitive, oubliant une nouvelle fois l'importance de l'enquête dans leur travail.

Les journalistes interviewés se disent inondés de communiqués de presse mais, en pratique, les utilisent largement pour choisir leurs sujets et rédiger leurs articles. Ils justifient leur recours à cette solution de facilité par manque de temps et par la possibilité qu'elle leur donne de consacrer par ailleurs plus de temps à des dossiers complexes. Cependant, les journalistes établissent une hiérarchie parmi les communiqués de presse. Les mieux considérés sont ceux des revues scientifiques prestigieuses. A l'opposé, ceux de l'industrie pharmaceutique et des sociétés de communication sont considérés avec beaucoup de méfiance. Entre les deux, les communiqués de presse rédigés par les institutions scientifiques sont considérés comme moins crédibles que ceux des revues prestigieuses et sont moins systématiquement lus en raison de leur nombre toujours croissant. Nous n'avons pas pu vraiment croiser ces commentaires avec le contenu des 426 articles analysés au chapitre III. Néanmoins nous avons noté que beaucoup d'articles semblaient rédigés par copier-coller des dépêches d'agences de presse (principalement *Associated Press* et *Reuters*). Le recours à une solution de facilité décrit par les interviewés semble donc assez général. Nous n'avons par

contre pas recherché une éventuelle similarité de contenu entre les dépêches d'agences de presse et les communiqués de presse dont elles pourraient être issues.

Notre enquête a aussi permis de révéler que pour les journalistes interviewés, les critères de nouveauté et de pertinence pour le public sont des critères essentiels, plus importants que la qualité et la fiabilité des résultats de la recherche. Plusieurs journalistes ont en effet constaté que ce n'était pas la science avérée et confirmée qui était couverte même si ils le regrettent. Selon eux, les résultats des études de réplication ne vont pas attirer les lecteurs et ne sont donc pas vendeurs. Sur ce point, les commentaires des interviewés sont en accord avec le contenu des articles de presse que nous avons examinés: les études portant sur les facteurs de risque liés au style de vie et les études initiales sont préférentiellement couvertes par la presse anglo-saxonne. D'autre part, les études contredisant des études déjà couvertes ne sont quasiment jamais reprises par la presse (chapitre III).

Le poids des routines dans la sélection des sujets et les contraintes de temps sont autant de facteurs qui biaisent la couverture des sciences et il semble que les journalistes en soient conscients. Il n'en reste pas moins que, pour la plupart des journalistes interviewés, ces problèmes sont inhérents au fonctionnement des médias. Pour eux, le traitement de l'actualité scientifique paraît une tâche moins prioritaire que la rédaction de dossiers plus étoffés sur des questions complexes

3.2.3- Le poids de la hiérarchie

Les conflits qui existent entre les journalistes et leurs rédacteurs en chef apparaissent clairement dans l'enquête. La sélection des sujets est influencée par les conférences de rédaction et les journalistes salariés ne choisissent leurs sujets que dans 63% des cas. Pour les pigistes, c'est encore moins : en moyenne 50% des sujets qu'ils proposent et sur lesquels ils ont travaillé sont acceptés. Des journalistes ont précisé que certains de leurs projets avaient été refusés parce qu'ils ne reflétaient pas l'opinion des rédacteurs en chef ou des directeurs du journal.

Les conflits et les différences de point de vue sont particulièrement visibles au niveau du choix des titres. Les journalistes ne choisissent pas les titres de leurs articles même si ceux que nous avons interviewés en proposent toujours un. Celui-ci n'est en général pas retenu et les journalistes sont souvent déçus par les titres choisis par la rédaction parce qu'ils les jugent exagérés et trompeurs. Lorsqu'ils s'en plaignent il est rare qu'ils aient gain de cause. S'attendant à des exagérations au niveau du titre, ils essayent de les contrebalancer par de meilleures explications dans le corps de l'article. C'est une situation que les journalistes déplorent mais contre laquelle ils n'ont pas beaucoup de prise.

« Il faut être réaliste sur la raison d'être des médias et sur les motivations de chaque personne aux différents niveaux. Et comme on le voit tout le temps ce qui compte ce n'est pas que se soit juste ou vrai, ça c'est une des motivations qui rentre en compte. Et bon, quand on y pense, ce que les médias veulent, c'est être lus. C'est intéresser, attirer le public, et au bout du compte, pour la presse écrite c'est combien de journaux, combien d'encarts publicitaires ils vont pouvoir vendre. » S9

Ce point de notre enquête éclaire une observation faite au chapitre III : si l'on compare tous les articles couvrant une même étude scientifique les titres sont très différents alors que les corps de textes paraissent plus homogènes. Chaque media cherche à se différencier de ses concurrents, ce qui pourrait expliquer le choix de titres distincts par les différentes rédactions. Par contre l'homogénéité des corps de textes pourrait résulter d'un « bâtonnage » des mêmes communiqués de presse ou dépêches d'agences.

3.3- Méconnaissance du fonctionnement de la recherche

La recherche avance lentement en partant de résultats initiaux prometteurs mais incertains pour arriver à un consensus après reproduction des résultats par d'autres équipes de recherche. La science est en construction et la validité d'une observation est donc relative, tout particulièrement dans le cas d'une observation initiale. Les journalistes interviewés dans l'enquête reconnaissent ce processus lent de la recherche et certains sont conscients de l'importance de la réplication mais tous semblent l'oublier quand ils rapportent les résultats de la recherche biomédicale. C'est comme si le fait d'être publié dans un journal prestigieux, par une équipe de recherche renommée ou d'obtenir le commentaire d'une sommité sur le sujet, donnait une crédibilité définitive au résultat. Comme nous l'avons montré au chapitre III, les journalistes sous-estiment ou ignorent délibérément le caractère incertain des observations qu'ils rapportent. En effet, plusieurs journalistes ont tenu des propos contradictoires en disant que les grandes revues scientifiques pouvaient se tromper et « monter des trucs en épingle » mais aussi dire, pour expliquer pourquoi ils n'ont pas eu recours à plusieurs sources ou pourquoi ils avaient « bâtonné la dépêche », « *c'est superflu, puisqu'il y a le peer-review qui intervient avant la publication de ce genre de papier* ».

Nous avons été surpris de constater que même les deux journalistes scientifiques docteurs en science, et avec une expérience post-doctorale en laboratoire, ne sont pas plus prudents que les autres dans la façon dont ils rapportent les résultats de la recherche. La formation scientifique, si elle leur permet sans doute de mieux appréhender certains sujets très techniques, ne parvient pas à leur faire éviter les écueils du journalisme : il faut de la nouveauté et du sensationnel. Evoquer l'incertitude inhérente à la recherche, rappeler la nécessité de confirmer les résultats n'est pas pour eux une priorité.

3.4- l'illusion du 4^{ème} pouvoir ?

Les médias sont souvent décrits comme le 4^{ème} pouvoir face aux pouvoirs politique, judiciaire et législatif. Ils se voient attribuer ce rôle de contre-pouvoir, notamment parce que le public, les spécialistes et les politiques pensent que ce sont eux qui font les nouvelles, décident de ce qui est important, mettent en lumière des problèmes ou qui, autrement dit, interviennent dans la mise à l'agenda (Neveu 2009). L'indépendance et l'objectivité journalistiques sont considérées comme les garants de ce contre-pouvoir. Ce pouvoir attribué aux médias incitent les scientifiques, leurs institutions et les éditeurs des journaux scientifiques à élaborer des stratégies leur permettant d'attirer leur l'attention afin d'accéder à la sphère publique. C'est ce que les auteurs allemands en sciences sociales ont décrit comme la *médialisation* (Peters 2012; Weingart 1997). Les universités et instituts de recherche se dotent donc de services de presse qui envoient des communiqués de presse aux journalistes (Peters et al. 2008) et les revues scientifiques prestigieuses mettent en valeur les articles qu'elles publient par des résumés grand-public, des commentaires et des éditoriaux (Entwistle 1995; Peters et al. 2008; Woloshin et al. 2002).

Comme nous l'avons vu, aussi bien au travers de l'analyse de la couverture médiatique des études scientifiques que dans l'enquête auprès des journalistes, cette démarche active des institutions scientifiques auprès des médias porte ses fruits. Il n'est en effet pas inhabituel que les journalistes reprennent ces résumés et communiqués de presse en les accompagnant éventuellement de citations de chercheurs. Les institutions de recherche influencent donc fortement la diffusion des sciences auprès du public. Les journalistes ne prennent qu'une part minime à cette activité de communication. Quelle que soit l'idée qu'ils s'en font dans l'idéal, ils font ici fi du scepticisme si cher à leur profession (Jensen 2010). Le rôle de la presse en tant que 4^{ème} pouvoir est donc quasiment inexistant lorsqu'il s'agit de couvrir les résultats de découvertes scientifiques. L'indépendance et l'objectivité journalistiques ne sont pas sollicitées dans ce contexte. Cette trop grande dépendance aux sources a déjà été remarquée et critiquée que ce soit dans la présentation des sciences ou des nouvelles en général dans la presse (Jensen 2010; Lewis et al. 2008; Neveu 2009; Watts 2014; Weitkamp 2014).

Ce constat global présente bien entendu des exceptions. Toutes les sciences ne sont pas couvertes de la même façon et pour certains sujets, en général controversés, les journalistes réalisent un réel travail d'enquête en interrogeant des sources avec d'autres points de vue (Champagne et al. 1994; Comby 2015; Friedman 2011; Marchetti 2010). Il n'en reste pas moins que, pour l'essentiel, le public n'est pas informé de l'incertitude inhérente aux résultats de la recherche biomédicale. Les journalistes que nous avons interviewés conçoivent leur rôle en tant qu'informateur du public ou de « critique du pouvoir ». Or nous avons pu

constater que dans la couverture médiatique des études scientifiques de notre base de données, l'information fournie au public est bien souvent une copie des dépêches d'agence sans analyse critique. Le prétendu rôle de « *gatekeeper* » des journalistes tient ici plus à l'idée qu'ils s'en font qu'aux des pratiques nécessaires à sa garantie.

En somme, l'enquête confirme que les journalistes ne prennent pas en compte l'incertitude scientifique lorsqu'ils couvrent les résultats de la recherche biomédicale. Il semblerait cependant que, bien souvent, ce soit parce qu'ils comprennent mal le fonctionnement de l'activité scientifique et qu'ils accordent une trop grande confiance au système du *peer-review* et aux contenus des journaux scientifiques prestigieux. Ils sont d'autant plus vulnérables à leurs sollicitations que leurs conditions de travail se sont détériorées et qu'ils disposent de temps courts pour rédiger leurs articles. D'autre part, ils indiquent aussi, en accord avec la littérature, que leurs relations avec les chercheurs se sont améliorées et que ceux-ci sont beaucoup plus ouverts. Ces observations sont à rapprocher du concept de médialisation et de l'évolution de l'activité scientifique vers un mode 2 où les interactions avec les médias sont valorisées et recherchées²⁸.

²⁸ Ces points sont plus largement discutés dans le chapitre suivant

CHAPITRE V

Conclusion

L'objet d'étude de cette thèse a été l'incertitude scientifique, telle qu'elle est décrite en sciences de la vie, et la façon dont elle est présentée au grand public via les médias. En effet, la production de connaissances se fait selon un processus incrémental de validation d'observations par réplique par les pairs. Un résultat initial est donc incertain et nécessite d'être reproduit pour être validé. Comme nous l'avons décrit dans l'introduction, cette définition de l'incertitude a peu intéressé les sciences sociales et les sciences de la communication. Cependant, l'existence de cette incertitude inhérente aux résultats initiaux de la recherche ne peut être ignorée ni des chercheurs en sciences de la vie, ni des journalistes et des décideurs. En effet, dans le chapitre II, nous avons confirmé et étoffé les observations faites sur les résultats de la recherche biomédicale : les résultats initiaux sont souvent exagérés et la majorité n'est pas confirmée par les études ultérieures sur le même sujet (Dumas-Mallet et al. 2016). C'est ce que les auteurs décrivent comme la « crise de la reproductibilité ». Nous avons également montré, dans le chapitre III, que la presse favorisait la couverture de résultats d'études initiales publiées dans des journaux prestigieux. Malheureusement, celles-ci ne sont pas plus souvent confirmées que les études publiées dans des journaux à comité de lecture, mais moins prestigieux. De plus, la presse anglo-saxonne ne se fait quasiment jamais l'écho des études ultérieures contredisant ces études initiales (Dumas-Mallet et al. 2017). Les résultats d'articles de recherche biomédicale présentés dans la presse sont donc exagérés voire faux et, par conséquent, le public au sens large est mal informé.

Nous avons également analysé le contenu de différents articles de presse couvrant des études initiales (Dumas-Mallet et al. 2017soumis). En effet, nous voulions étudier la façon dont les journalistes et leurs rédacteurs en chef « cadraient » les résultats de ces recherches initiales. Cette analyse a montré que les journalistes décrivaient l'étude comme initiale dans plus de la moitié des articles mais que, par contre, la nécessaire validation de ces résultats était peu mentionnée (21% des cas). D'autre part, cette mention était principalement le fait des scientifiques, qu'ils soient les auteurs de ces études ou bien des experts interrogés par les journalistes. Nous avons aussi montré, dans le chapitre III, que cette mention tendait à disparaître dans les articles de presse les plus récents alors que la reproductibilité des résultats n'était cependant pas meilleure et qu'elle faisait l'objet de nombreux articles dans les revues scientifiques.

Enfin, dans une enquête réalisée auprès de 21 journalistes scientifiques décrite au chapitre IV, nous avons pu montrer que ceux-ci utilisaient beaucoup les communiqués de presse des journaux scientifiques prestigieux, qu'ils considèrent comme des sources fiables d'information. L'enquête a révélé que près d'un tiers des journalistes semble connaître le processus de production de connaissances par validation des résultats grâce à leur réplique par d'autres équipes de chercheurs. Néanmoins, dans bien des cas, la réserve qu'ils devraient témoigner à l'égard de résultats initiaux disparaît lorsque ces résultats sont publiés dans des journaux prestigieux. Cela donne aux résultats initiaux un crédit immédiat.

Plus généralement, les analyses que nous avons réalisées au cours des différents chapitres de cette thèse, ainsi que la littérature existante sur ce sujet (Alberts 2013; Brembs et al. 2013; Callahan et al. 2002; Casadevall et al. 2014b; Reich 2013; Simons 2008) ont montré l'importance des journaux scientifiques à fort facteur d'impact dans l'activité scientifique. Comme nous le verrons tout au long de ce chapitre, ces journaux semblent être considérés par les différents acteurs - scientifiques, journalistes et décideurs - comme les sources principales, voire uniques, d'avancées scientifiques. Nous nous efforcerons de montrer, au cours des trois parties de ce qui suit, les avantages que représente la publication dans des journaux prestigieux, tant au niveau de l'activité scientifique que de l'évaluation de la recherche ou encore de son impact sur le grand public. Tout d'abord, nous rappellerons en quoi l'activité scientifique, telle qu'elle est pratiquée et envisagée par les chercheurs en sciences biomédicales, est attachée aux normes mertonniennes (Merton 1973) et donc à un mode 1 de production des connaissances où l'activité scientifique est gérée et orientée par la communauté scientifique. L'autonomie de la science en est l'attribut central selon les chercheurs en sciences biologiques (Ziman 1996, 1996b). Puis nous expliquerons que les conditions actuelles d'évaluation et de financement de la recherche poussent les scientifiques à employer des pratiques de recherche discutables et à se rattacher plus à des contre-normes telles que celles de quantité et d'intérêt personnel (Anderson et al. 2010; Mitroff 1976). Nous verrons que des critères extérieurs à la production de connaissances (financement de la recherche et visibilité médiatique) interviennent lourdement dans la pratique de l'activité scientifique et, surtout, dans son approche contemporaine de la reproductibilité. Enfin, au travers des résultats que nous avons obtenus sur la couverture médiatique de la recherche biomédicale et de l'enquête réalisée auprès des journalistes scientifiques, nous montrerons que l'activité de recherche dans le domaine biomédical présente certains attributs du mode 2 ou 2.0 (Bucchi 2015; Gibbons et al. 1994) dans lequel l'interaction avec les médias est primordiale et valorisée par les scientifiques (Peters 2012b).

I- Crise de la reproductibilité et journaux scientifiques à fort facteur d'impact

La crise de la reproductibilité est un problème familier des chercheurs en sciences biologiques et médicales et les préoccupe beaucoup (Chalmers et al. 2014; Ebersole et al. 2016; Ioannidis 2014c; Macleod et al. 2014; Nosek et al. 2012) comme en témoigne l'explosion du nombre de références sur PubMed (voir Chapitre III). La norme de scepticisme organisé décrite par Merton explique cette attitude : les scientifiques se doivent d'être critiques vis à vis de tout énoncé, le résultat doit être analysé, vérifié puis confirmé. La reproductibilité est donc ici un attribut central de l'activité scientifique.

Cependant, la non-reproductibilité des résultats n'est pas, à notre avis, le problème principal. En effet, comme plusieurs auteurs le défendent, il serait illusoire de vouloir obtenir 100% de reproductibilité (Open Science Collaboration 2015) et « les publications scientifiques ne sauraient être considérées comme des faits établis de vérités indiscutables ; elles représentent plutôt de petits pas prometteurs au travers de la jungle de l'ignorance » (Ziman 1969). La science est par nature incertaine et ne peut être validée que par réplication des résultats. Que tous les résultats ne puissent pas être reproduits n'est pas problématique. Le problème vient plutôt de la publication sélective de résultats positifs, présentés comme « extraordinaires » dans des journaux prestigieux et de la non-publication des résultats négatifs (Fanelli 2010b, 2012; Vinkers et al. 2015; Young et al. 2008). En effet, les journaux prestigieux tels que *Nature* et *Science* ont tendance à publier des résultats surprenants, novateurs et susceptibles d'attirer l'attention des médias (Bucchi 2015; Chapman et al. 2007; Chew et al. 2007; Franzen 2012).

Il n'en reste pas moins que publier dans une revue prestigieuse présente de nombreux avantages pour les auteurs : ils intègrent le « *Golden Club* » (Reich 2013). Ils acquièrent un capital de crédibilité et de réussite. C'est ce que Merton décrivait comme « l'effet Saint Matthieu » : la publication dans une revue prestigieuse accorde un crédit important au chercheur ce qui facilite l'obtention de financements et donc la publication de nouveaux résultats dans ces mêmes revues prestigieuses. Leur travail est reconnu par les pairs et est remarqué puisque non seulement les articles des revues prestigieuses sont plus visibles pour la communauté scientifique mais aussi pour des médias. En effet, des auteurs ont montré que les articles publiés dans des journaux à fort facteur d'impact sont plus cités (Barabasi et al. 2012; Jannot et al. 2013; Kulkarni et al. 2007; Larivière et al. 2014) et que la couverture médiatique de ces publications augmentait encore leur probabilité d'être citées (Phillips et al. 1991). Enfin, les articles publiés dans ces journaux sont plus souvent relayés par les médias (Dumas-Mallet et al. 2017; Gonon et al. 2012). Les avantages que se voient conférer les scientifiques lorsqu'ils arrivent à publier dans des journaux prestigieux sont donc notables. Ceci vient

illustrer l'apparition d'une contre-norme (Mitroff 1976) dans le comportement d'un grand nombre de chercheurs : ils n'agissent plus de manière désintéressée mais dans leur propre intérêt. De même, il semblerait que la norme d'universalisme soit également altérée puisque la communauté scientifique accorde un crédit plus important aux résultats publiés dans des journaux prestigieux que dans les autres journaux à comité de lecture (Simons 2008). Ainsi dans une analyse du nombre de citations reçu par des articles publiés dans 5 journaux prestigieux (*Cell*, *Nature*, *Science*, *Proceedings of the National Academy of Science* et *New England Journal of Medicine*), Barabasi et al ont montré que le plus haut centile (1%) des articles les plus cités de chacun de ces journaux totalisait 17% des citations en 2010 (Barabasi et al. 2012). Par conséquent, un petit nombre d'articles scientifiques publiés dans des journaux prestigieux domine les citations.

Cette prépondérance des journaux prestigieux s'est d'autre part accentuée. En effet, de l'accord de plusieurs auteurs, il y a eu une évolution de la pratique de l'activité scientifique (Bonneuil 2004; Dubois 2013; Elzinga 1997; Pellizzoni 2003). Cette évolution peut être expliquée par les politiques d'évaluation de la recherche (Osterloh et al. 2008). En effet, Osterloh et al, mentionnent « les dérives entraînées par les indicateurs numériques de production » et de l'obligation pour les chercheurs de « rendre des comptes ». Cette « rentabilité » nécessaire de la recherche pousserait donc les chercheurs à orienter leur activité vers des recherches publiables, susceptibles d'être bien évaluées, notamment si celles-ci sont publiées dans des journaux prestigieux (Osterloh et al. 2008).

II- Politique d'évaluation de la recherche, financements et journaux à fort facteur d'impact

De nombreux chercheurs en sciences biomédicales attribuent en partie ce problème de reproductibilité à l'hypercompétitivité qui existe désormais pour le financement de la recherche pour une communauté scientifique toujours plus importante (Alberts et al. 2014). Par exemple, on estime que le budget du NIH (*National Institutes of Health*) est désormais de 25% inférieur, en dollars constants, à celui qu'il était en 2003 (Alberts et al. 2014). Par conséquent, pour faciliter les choix des instances de financement de la recherche, les chercheurs sont évalués sur des critères aisés d'utilisation et compris par des chercheurs de différentes disciplines : le nombre et le prestige de leurs publications. Ces indices numériques, devenus primordiaux pour l'évolution de la carrière et l'obtention de financements, sont sévèrement critiqués par les scientifiques (Alberts 2013; Casadevall et al. 2014b; Edwards et al. 2017; Ioannidis 2011; Lawrence 2003) et certaines institutions (Comité d'éthique du CNRS et al. 2014). Ainsi, le comité d'éthique du CNRS (COMETS) dénonce les dérives de la

politique d'excellence qui utilise des critères empruntés au management des entreprises privées comme critères d'évaluation de la recherche (Comité d'éthique du CNRS et al. 2014).

Une enquête, publiée dans la revue *Nature*, évalue la perception et l'influence des indicateurs numériques sur la carrière des scientifiques du point de vue des chercheurs eux-mêmes, mais aussi de ceux qui les recrutent (Abbott et al. 2010). Cette enquête réalisée auprès de 150 de lecteurs révèle que 75% des scientifiques interrogés pensent que le nombre et le facteur d'impact des journaux dans lesquels ils ont publié ont joué un rôle majeur dans leur recrutement et leur évolution de carrière. Ainsi il paraît vraisemblable, comme le souligne le comité d'éthique du CNRS (COMETS), que ces critères incitent certains chercheurs à se livrer à des pratiques pour le moins discutables et « non conformes à l'éthique » afin de publier dans des revues prestigieuses (Alberts et al. 2014; Casadevall et al. 2014b; Comité d'éthique du CNRS et al. 2014), ils préfèrent soumettre leurs travaux à des journaux avec de fort impact facteur même si ceux-ci sont moins spécialisés (Fraser et al. 2009). D'autre part, dans l'enquête réalisée par la revue *Nature*, 51% des chercheurs interrogés déclarent qu'ils ont modifié leur comportement en réponse à ce système d'évaluation. En somme, les incitations à publier beaucoup, régulièrement et de préférence dans des journaux prestigieux conduisent les scientifiques à poursuivre des objectifs de recherche faciles à atteindre et donc publiables, à éviter les sujets « risqués » et à ne pas confirmer les résultats déjà obtenus (Alberts et al. 2014; Comité d'éthique du CNRS et al. 2014). En effet, être la deuxième équipe à mettre en évidence un phénomène n'a pas beaucoup de valeur aux yeux des éditeurs des journaux scientifiques. A ceci s'ajoutent les pertes de temps considérables des scientifiques, aussi bien ceux qui rédigent les demandes de financements que ceux qui les évaluent (Alberts et al. 2014).

C'est ce que décrit Bucchi quand il explique l'évolution de la production de connaissances vers un mode 2.0 (ou mode 2) : la compétition pour les ressources telles que le financement ou un poste académique modifie l'activité scientifique elle-même (Bucchi 2015). Cette évolution est encore plus flagrante pour la recherche biomédicale : en effet, c'est une recherche par définition appliquée et dont les enjeux pour la santé sont importants. Cette recherche intéresse donc tout naturellement les scientifiques, les décideurs (on peut citer ici le défi « Santé et Bien-être » proposé en France par la Stratégie Nationale de Recherche) et le public. On voit en quoi ce type de recherche se rattache à un mode 2 de production de connaissances.

Premièrement, la recherche est orientée vers des applications directes, comme le montre le cadre des appels à projet ou la politique de recherche. Par exemple, l'orientation 18 du défi santé-bien-être de la stratégie nationale de recherche est « d'augmenter la qualité et

l'attractivité de la recherche clinique, le nombre d'essais réalisés en France » (Ministère de l'Education Nationale 2015). On peut également citer « l'appel à projets de recherche internationaux sur les maladies neurodégénératives : approches innovantes III » proposé par l'agence nationale de la recherche (ANR). Elle précise que « les projets pourront inclure des études destinées (...) à créer des avancées technologiques visant à améliorer le diagnostic ou le traitement des maladies neurodégénératives ».

Deuxièmement, et en lien avec le point précédent, la recherche scientifique de nos jours est principalement financée sur projets : les chercheurs répondent à des appels d'offre sur des sujets « à la mode » susceptibles d'être financés (Hanna 2015). Tout un pan de la littérature biomédicale critique d'ailleurs vivement cette évolution du mode de financement de la recherche. En effet, la rédaction des dossiers de candidature prend un temps et une énergie considérables au détriment de la pratique de la recherche (Ioannidis 2011). Leur évaluation favorise des sujets aptes à donner rapidement des résultats pas trop ambitieux (Lawrence 2003) et les chercheurs ont bien souvent tendance à promettre plus qu'ils ne pourront jamais obtenir (Jouvenet 2011; Lawrence 2003, 2009). La sociologie de la promesse permet d'expliquer en partie le comportement de ces chercheurs. En effet, notamment dans la recherche biomédicale, les résultats attendus peuvent avoir des implications importantes positives pour les patients (identification de facteurs de risque, outils diagnostic, traitements). Elle reste cependant une recherche couteuse et soumise à appel à projets. Les chercheurs, afin de sécuriser des financements peuvent donc « adapter » leurs projets de manière à ce qu'ils « cadrent » avec les attentes des agences de financements. Ils promettent donc des résultats sur la base d'expériences à venir (Chapitre I).

Troisièmement, les sciences de la communication ont montré que les scientifiques se préoccupaient de l'opinion du public et interagissaient volontiers avec les journalistes « pour éduquer le public et donner une opinion favorable aux sciences » (Peters et al. 2008). En effet, plusieurs enquêtes ont montré qu'une des motivations principales des scientifiques d'interagir avec les médias était d'améliorer l'opinion du public vis à vis des sciences en l'éduquant (Besley et al. 2013; de Semir 2010; Nisbet et al. 2002b). Le rationnel étant que, si le public connaissait et comprenait mieux les sciences, il y serait plus favorable. Ce point a d'ailleurs été largement critiqué par différents auteurs notamment en raison de son caractère normatif (Schafer 2009). D'autre part les institutions scientifiques (i.e. organismes de recherche, agences de financements) encouragent depuis quelques années la diffusion de l'information scientifique vers le public ; ce qui indique là encore l'importance accordée à la légitimation par le public (Lettre de l'Institut National des Sciences Biologiques 2016; The Royal Society 2006). Ainsi, le comité d'éthique du CNRS (COMETS) encourage les scientifiques à

divulguer les sciences, à « ne pas sous-estimer les réticences de la population à l'égard de la science » et de préciser que « le temps passé hors de leur communauté par ceux qui souhaitent s'impliquer dans la communication scientifique, pris sur celui de la recherche, doit être valorisé dans l'évaluation de la carrière » (Comité d'éthique du CNRS et al. 2014). Ceci dénote une évolution de l'opinion des institutions scientifiques et des scientifiques eux-mêmes vis à vis de leurs rôles sociaux et de leurs devoirs en ce qui concerne la société. Cette évolution est aussi visible au niveau de leurs attitudes vis à vis des médias (Peters et al. 2008b).

Ces 3 caractéristiques – orientation vers les applications, financements sur projet et interactions grandissantes avec les médias et donc plus largement la société – sont similaires à trois des attributs du mode 2 (Chapitre I). Si l'existence d'un mode 1 ou 2 de production de connaissances est largement critiquée par les sociologues des sciences (Shinn 2002), on voit que pour les chercheurs en sciences biomédicales cette évolution et ces changements de la pratique de l'activité scientifique sont perçus comme un passage à un mode 2. Autrement dit, précédemment, ils ressentaient moins ces impératifs de publication, de communication ou la compétition pour les financements.

III- La visibilité médiatique et journaux à fort facteur d'impact

Ce qui souligne sans doute le plus le changement de production de connaissances c'est l'orientation grandissante de l'intérêt des scientifiques et de leurs institutions vers les médias « grand public » ou généralistes. C'est ce que plusieurs auteurs en sciences sociales décrivent comme la *médialisation*, ou, autrement dit, l'orientation des scientifiques vers les médias en raison de la capacité qu'ils leur accordent de façonner l'opinion et de mettre en lumière les thèmes qui nécessitent l'attention des décideurs (Peters 2012; Weingart 1997). Ceci explique que les scientifiques, leurs institutions et les éditeurs de journaux scientifiques essaient d'influencer le contenu et la façon dont les sciences sont décrites dans les médias (Autzen 2014; Franzen 2012; Peters et al. 2008b). Cette orientation vers les médias induit donc un changement dans la façon dont la science est conduite : les scientifiques incorporent des logiques médiatiques dans leur pratique de la recherche (Bauer 2012). Plusieurs observations viennent étayer cette orientation croissante vers les médias. Tout d'abord, de nombreux auteurs s'accordent à dire que les relations entre les scientifiques et les journalistes ont évolué et que les deux parties sont majoritairement satisfaites de leurs interactions. Ainsi, selon une enquête réalisée auprès de 1354 chercheurs originaires de 5 pays (France, Allemagne, Japon, Royaume-Uni et Etats-Unis), 75% d'entre eux étaient satisfaits de leur interaction avec les médias et pensaient que la couverture médiatique de leurs travaux était bénéfique pour leur

carrière. Seulement 3% des chercheurs interviewés estimaient que cette couverture médiatique avait des conséquences principalement négatives (Peters et al. 2008b). Les chercheurs sont aussi moins critiques sur la façon dont leurs résultats sont présentés et sont prêts à accepter des simplifications et même des inexactitudes (Gascoigne et al. 1997; Peters 2012b). Les résultats de notre enquête auprès des journalistes, décrits au chapitre IV, confirment cette amélioration : plusieurs ont mentionné une évolution positive de leurs rapports avec les scientifiques (i.e. les scientifiques sont plus ouverts et plus disponibles).

Ensuite, les institutions de recherche se sont dotées de services de presse. Ceux-ci rédigent des communiqués de presse à l'attention des journalistes. Ils y opèrent une sélection des données importantes et « formatent » leur texte en utilisant des critères d'intérêt médiatique (Peters et al. 2008b; Samuel et al. 2017; Schafer 2011b). Les éditeurs des journaux scientifiques prestigieux proposent eux aussi des communiqués de presse aux journalistes dont certains sont soumis à embargo. Ces stratégies (communiqué de presse et/ou embargo) ont pour conséquence de favoriser la couverture médiatique de ces études : en effet, les journalistes vont avoir tendance à s'en faire l'écho car ils pensent que leurs concurrents risquent de le faire (Franzen 2012). Ainsi, les scientifiques essaient de tirer avantage de leur visibilité médiatique pour pouvoir orienter la recherche ou en tout cas contrôler la recherche qui est présentée dans les médias (Peters et al. 2008b). Certains éléments de notre analyse du contenu des articles de presse viennent étayer cette observation. Comme nous l'avons vu au chapitre III, la plupart des articles de presse couvrant une même publication scientifique sont assez similaires, ce qui suggère que les journalistes ont vraisemblablement tous utilisé la même source.

En somme, le fait de publier dans un journal prestigieux augmente *de facto* la visibilité médiatique du chercheur. La plupart des scientifiques ne recherchent sans doute pas en premier lieu l'attention des médias ; elle semble plutôt être une conséquence indirecte de la publication de leurs articles dans des journaux prestigieux. Par contre, des enquêtes auprès des éditeurs des journaux prestigieux montrent que certains d'entre eux recherchent et encouragent l'attention des médias, non seulement dans le choix des articles qu'ils publient mais aussi lorsqu'ils accompagnent ces articles de communiqués de presse ou d'éditoriaux (Chew et al. 2007). Certains, dans le but d'augmenter le facteur d'impact de leur journal, se prêteront même à des manquements flagrants à l'éthique éditoriale en insistant pour que les auteurs qu'ils publient citent principalement des études publiées dans leurs pages (Casadevall et al. 2014b). L'importance que les éditeurs accordent aux médias et donc à des critères extra-scientifiques est largement critiquée. Ainsi, Randy Schekman, prix Nobel de médecine a écrit

une tribune à ce sujet dans *The Guardian* où il accuse les journaux *Science*, *Nature* et *Cell* de nuire à l'activité scientifique (Schekman 2013). Comme le souligne également cyniquement Richard Smith, ancien éditeur du journal *BMJ* (*British Medical Journal*) « les éditeurs peuvent critiquer les médias, ils adorent qu'ils s'intéressent à leur journal ». Les éditeurs des journaux médicaux (i.e. *the Lancet*, *the BMJ* ou *the NEJM*) proposent et envoient des communiqués de presse aux journalistes car la couverture médiatique des articles qu'ils publient non seulement augmente leur facteur d'impact mais leur permet aussi de toucher une plus grande part de leur lectorat potentiel, notamment les médecins (Smith 2006).

Cette visibilité médiatique contribue également à augmenter le prestige des chercheurs. Elle leur donne une bonne réputation et suggère une excellence de la qualité de la recherche et donc une crédibilité importante (Bucchi 2015).

On voit alors que la politique d'évaluation de la recherche, sur la base des facteurs d'impact dans lesquels les articles sont publiés, est dangereuse pour l'activité de recherche elle-même, mais pour le public et les décideurs également. Pour l'activité de recherche puisque des études ont montré que le nombre de citations d'un article dépendait plus du facteur d'impact du journal dans lequel l'étude avait été publiée que de la qualité de cette étude (Callahan et al. 2002) et que des erreurs y étaient plus souvent soulignées. D'autre part, ces études ne sont pas plus valides (Dumas-Mallet et al. 2016) et la qualité de la recherche (i.e. description de la méthodologie et analyse statistique) n'est pas meilleure (Franzen 2012). Ainsi, des chercheurs peuvent alors s'engager dans des recherches basées sur des résultats erronés entraînant donc des pertes de temps et d'argent (Kivimaki et al. 2014). D'autre part, la volonté de publier dans un journal à fort impact peut ralentir le processus de la communication scientifique : en effet en essayant d'être publiés dans de tels journaux les chercheurs passent plus de temps à faire des expériences complémentaires pour être acceptés ou à soumettre leurs résultats en chaîne à d'autres journaux lorsque les premiers éditeurs les rejettent (Casadevall et al. 2014b).

Enfin, certains scientifiques ont tendance à choisir des critères d'intérêt médiatique lors de la rédaction de leur article pour des journaux prestigieux (Fraser et al. 2009) et à exagérer leurs résultats dans leurs interactions avec les médias (Holtzman et al. 2005). Ainsi plusieurs études ont montré que, si les résultats étaient souvent exagérés dans les médias que ce soit la presse, la télévision ou internet (Haneef et al. 2015; Holtzman et al. 2005; Woloshin et al. 2009; Yavchitz et al. 2012), ceci était dû à des exagérations déjà présentes dans les communiqués de presse rédigés par les institutions scientifiques (Sumner et al. 2014; Woloshin et al. 2009 b), les journaux scientifiques (Sumner et al. 2016; Woloshin et al. 2002)

et même au sein des publications (Gonon et al. 2011; Woloshin et al. 2009). Ces exemples montrent l'intérêt des scientifiques et de leurs institutions pour le contrôle de la diffusion des résultats scientifiques et notamment pour sa légitimation (voir paragraphe suivant). Cette stratégie est par ailleurs efficace puisque la présence d'un communiqué de presse augmente énormément la probabilité pour une étude d'être reprise dans la presse et sur internet (Bartlett et al. 2002; de Semir et al. 1998; Haneef et al. 2017).

Les conséquences négatives sur le public et les décideurs sont aussi importantes. En effet, les articles relayés par les médias sont souvent les résultats d'études initiales peu fiables (Dumas-Mallet et al. 2017; Gonon et al. 2012). De plus, le contenu des articles de presse que nous avons analysé indique que l'incertitude liée à ces résultats est peu mentionnée et que les mentions tendent même à disparaître (Dumas-Mallet et al. 2017). L'information scientifique présentée dans la presse est donc biaisée, voire même fautive. Ceci est problématique puisqu'une enquête auprès de plusieurs décideurs a montré que les articles de presse influencent leur décision (Petersen et al. 2010). Les auteurs de cette enquête révèlent que les décideurs politiques (dans leur cas 39 décideurs politiques allemands responsables de sujets en biomédecine et particulièrement en santé et en bioéthique) utilisent les informations sur les découvertes et développements scientifiques présentés dans les médias, non seulement pour se former une opinion mais aussi pour déterminer ce qui constitue les nouveaux « défis » (Petersen et al. 2010). Les décideurs basent donc certaines de leurs opinions sur des résultats non vérifiés voire même faux. Ceci est d'autant plus préoccupant que les médias ne font que rarement état des réfutations de résultats scientifiques qu'ils ont pourtant relayés (Dumas-Mallet et al. 2017). Ainsi, la justification du traitement du TDAH par un psychostimulant est elle encore basée sur une information scientifique qui a été par la suite contredite de nombreuses fois mais dont ni la presse, ni les scientifiques et finalement les décideurs ne semblent faire cas (Gonon 2009; Gonon et al. 2014). D'autre part, plusieurs études ont montré les conséquences négatives de la couverture de certains résultats initiaux sur la population. Ainsi, la couverture médiatique intensive de nouveaux traitements potentiels a non seulement des conséquences sur les demandes et comportements des patients, mais aussi sur la pratique de la médecine (Brunt et al. 2003; Grilli et al. 2002; Marshall 1998; Sharma et al. 2003).

Plusieurs auteurs appellent donc les journalistes à être plus critiques envers les sciences qu'ils couvrent et notamment lorsqu'ils rapportent les résultats de publications scientifiques (Dentzer 2009; Oransky 2015; Watts 2014). Selon ces auteurs, l'attitude des journalistes est dommageable, pour l'entreprise journalistique mais aussi pour le public et montre que les institutions scientifiques sont efficaces dans leur orientation de ce qu'elles jugent important à couvrir (Peters 2012). Notre enquête auprès des journalistes renforce

malheureusement ces observations : ils n'ont pas une attitude assez critique vis à vis des informations qu'ils reçoivent. En effet, ils ont confirmé qu'ils accordaient une grande confiance aux communiqués de presse des journaux scientifiques prestigieux. Cependant, certains deviennent plus prudents face à ceux des institutions scientifiques (CNRS ou INSERM) dont ils reconnaissent les stratégies d'auto-promotion (Chapitre IV). Il semblerait donc qu'il y ait une évolution des représentations des journalistes, comme l'indiquent les résultats de notre enquête mais aussi l'apparition d'articles ou d'ouvrages sur les problèmes de reproductibilité dans la presse grand public (Harris 2017; The Economist 2012, 2013a, 2013b). Ceci est encourageant et pourrait permettre une évolution de la pratique des journalistes, au moins lorsqu'ils rapportent les résultats de publications scientifiques.

IV- Vers un nouveau regard sur les rapports sciences-médias

Que cela soit en sociologie des sciences ou en sciences de la communication, plusieurs auteurs décrivent la modification de l'activité scientifique et l'intervention de facteurs extérieurs dans sa pratique (Bauer 2012; Bucchi 2015). Ainsi les politiques d'évaluation de la recherche, l'hypercompétitivité pour des ressources financières moins importantes ont conduit les scientifiques à modifier leur comportement : ils cherchent davantage à publier dans des journaux prestigieux et sont sensibles à l'intérêt médiatique que peut susciter leur recherche. Ces deux critères (publication dans des journaux prestigieux et couverture médiatique) leurs permettent d'obtenir financements et avancées dans leur carrière.

Ceci a des conséquences négatives pour la recherche biomédicale. En effet, le financement sur projet favorise parfois des projets de recherches sur des pathologies dont la prévalence est faible mais pour lesquelles il existe des financements alors que d'autres pathologies, avec des prévalences fortes sont ignorées (Hanna 2015). L'orientation des chercheurs en sciences biomédicales vers les médias, dans le but d'influencer la décision publique et le financement de la recherche, conduit certains à utiliser des critères susceptibles d'attirer leur attention. Ils choisissent des sujets porteurs, utilisent un vocabulaire sensationnaliste et ont tendance à exagérer la portée de leurs résultats (Holtzman et al. 2005; Nelkin 1996). L'incertitude est presque gommée et les résultats sont présentés dans les médias comme des faits. D'autre part, le système d'autocorrection de la recherche par réplication des résultats par les pairs est débordé (Edwards et al. 2017) : non seulement il n'y a pas d'incitation à reproduire les résultats mais lorsque des études de réplication sont faites, les résultats sont peu reproductibles (Begley et al. 2012; Ioannidis 2005b; Prinz et al. 2011).

Des résultats faux ou exagérés polluent les publications scientifiques, sont à l'origine de dogmes infondés ou de recherches inutiles (Begley et al. 2012; Greenberg 2009; Tatsioni et al. 2007). Il semblerait que les chercheurs en sciences biomédicales aient oublié que « l'information contenue dans un article scientifique primaire a à peu près autant de sens qu'un numéro de téléphone d'une page de l'annuaire : elle ne devient importante que lorsqu'elle est utilisée par d'autres pour être intégrée dans un schéma plus large » (Ziman 1969).

La confiance et le soutien du public pourraient eux aussi être minés. En effet, l'importance accordée à la visibilité médiatique plutôt qu'à sa qualité pose le problème de la crédibilité de la science (Peters et al. 2008b). Cette crédibilité et la capacité de la science à produire des connaissances fiables et aptes à permettre la décision sont au cœur des désaccords entre les sociologues des sciences. Ici nous souhaitons souligner le travail de Collins dans sa présentation d'une troisième « vague de *science studies* » (Collins et al. 2002). Nous le rejoignons notamment dans sa description des deux premières « vagues » de la sociologie des sciences. La première « vague » n'envisageait l'activité scientifique que comme productrice de vérités et en décrivait les normes, tandis que la deuxième, plus sceptique, s'est attachée à montrer que la science et les scientifiques n'étaient pas les seules sources de réponses ni de « vérités ». Selon Collins, cette deuxième vague s'est amplifiée au point que certains de ces sociologues ont considéré les sciences comme une forme de religion ou de politique (Collins 2009). Déplorant cet état de fait, Collins propose donc une troisième « vague » qui redonnerait à la science sa place tout en tenant compte de ses limites. Il donne une place particulière à l'expertise et à la mesure et à la compréhension du degré de consensus entre les experts lors de la présentation des sciences au public et aux décideurs politiques (Collins et al. 2017). Ceci rejoint ainsi notre argumentation en faveur d'une attention particulière des chercheurs et des journalistes lorsqu'ils rapportent des résultats de la recherche biomédicale : les chercheurs devraient rester prudents quant à la validité de leurs résultats et les journalistes devraient adopter une attitude critique vis à vis des résultats de la recherche. En effet, présenter les résultats d'études initiales comme certains, occulte et fausse aux yeux du public le fonctionnement normal de la recherche. Même si les récentes enquêtes au niveau européen montrent que le public est globalement satisfait de la façon dont les sciences sont présentées dans les médias (Eurobaromètre 2007), il ne faudrait sans doute pas considérer cette attitude pour acquise. D'autre part, et contrairement aux idées reçues, les exagérations pratiquées par les chercheurs ou leurs institutions sont inutiles puisque quelques enquêtes auprès du public ont montré que les résultats de la recherche biomédicale sont

d'autant plus crédibles que les scientifiques évoquent leurs doutes (Jensen et al. 2012; Jensen et al. 2013).

De plus, une enquête récente semble indiquer que la réputation des scientifiques ne serait pas entachée par une invalidation de certains de leurs résultats (Ebersole et al. 2016). Ce qui est le plus important pour la réputation des chercheurs aux yeux des personnes interrogées, qu'elles soient issues du public au sens large ou de la communauté scientifique, ce n'est pas que les résultats soient valides mais que, si ils ne le sont pas, les scientifiques le reconnaissent ou même le démontrent (Ebersole et al. 2016). Les scientifiques et leurs institutions devraient donc accorder une attention toute particulière à la façon dont ils communiquent les résultats au public. Ils devraient s'abstenir d'exagérer ou d'extrapoler les résultats. Enfin, il serait important d'évaluer, grâce à une enquête auprès de scientifiques, les conséquences du financement principalement sur projet de la recherche biomédicale. En effet, dans une enquête réalisée auprès de scientifiques travaillant dans les nanosciences, Jouvenet avait montré que ces changements étaient critiqués par les scientifiques. Ceux-ci reprochaient notamment les pertes de temps importantes ainsi que le manque de liberté pour le choix et la conduite de leurs projets (Jouvenet 2011). Le même type d'enquête pourrait être réalisé auprès de chercheurs en sciences biomédicales qui ont l'habitude ou non d'apparaître dans les médias. Dans cette optique, nous avons mis au point un questionnaire, présenté dans l'annexe 5, destiné aux scientifiques impliqués dans la recherche biomédicale afin de recueillir leur opinion sur une série de mesures proposées par des leaders de cette recherche dans le but d'améliorer la reproductibilité. Nous avons testé ce questionnaire auprès de chercheurs présents au colloque organisé par l'université de Bordeaux en janvier 2016: « l'intégrité scientifique, parlons-en ! ». L'analyse de cette enquête préliminaire nous a suggéré des modifications pour améliorer notre questionnaire. Nous le proposerons à d'autres chercheurs afin de recueillir leur opinion et de mieux comprendre leurs pratiques. Ce type d'enquête, ainsi que les leçons que l'on peut tirer de l'ensemble de cette thèse, pourrait ainsi, *a minima*, permettre de proposer des mesures scientifiquement fondées visant à améliorer la pratique de l'activité de recherche.

BIBLIOGRAPHIE

- Abbott, A., et al. (2010), 'Metrics: Do metrics matter?', *Nature*, 465 (7300), 860-2.
- Academy of Medical Sciences, et al. (2015), 'Reproducibility and reliability of biomedical research: improving research practice', (The academy of medical science).
- Alberts, B. (2013), 'Impact factor distortions', *Science*, 340 (6134), 787.
- Alberts, B., et al. (2014), 'Rescuing US biomedical research from its systemic flaws', *Proc Natl Acad Sci U S A*, 111 (16), 5773-7.
- Allgaier, J., et al. (2013), 'Medialized science? Neuroscientists' reflections on their role as journalistic sources.', *Journalism Practice*, 7 (0-4), 413-29.
- Anderson, M. S., Martinson, B. C., and De Vries, R. (2007), 'Normative dissonance in science: results from a national survey of u.s. Scientists', *J Empir Res Hum Res Ethics*, 2 (4), 3-14.
- Anderson, M. S., et al. (2010), 'Extending the Mertonian Norms: Scientists' Subscription to Norms of Research', *Journal of Higher Education*, 81 (3), 366-93.
- Angell, M. and Kassirer, J. P. (1994), 'Clinical research--what should the public believe?', *N Engl J Med*, 331 (3), 189-90.
- Ankney, R. N., Heilman, P., and Kolff, J. (1996), 'Newspaper Coverage of the Coronary Artery Bypass Grafting Report', *Sci Commun*, 18 (2), 153-64.
- Autzen, C. (2014), 'Press releases — the new trend in science communication', *Journal of Science communication*, 13 (03).
- Babbage, C. (1830), *Reflections on the decline of science in England and on some of its causes*. The Works of Charles Babbage (London Pickering: Campbell-Kelly M).
- Badenschier, F. and Wormer, H. (2012), 'Issue Selection in Science Journalism: Towards a Special Theory of News Values for Science News?', *The sciences' media connection-Public communication and its repercussions* (28: Springer Netherlands).
- Baker, M. (2016), 'Statisticians issue warning over misuse of P values', *Nature*, 531 (7593), 151.
- (2016b), '1,500 scientists lift the lid on reproducibility', *Nature*, 533 (7604), 452-4.
- Barabasi, A. L., Song, C., and Wang, D. (2012), 'Publishing: Handful of papers dominates citation', *Nature*, 491 (7422), 40.
- Bartlett, C., Sterne, J., and Egger, M. (2002), 'What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers', *BMJ*, 325 (7355), 81-4.
- Bauer, M. (1998), 'The medicalization of science news - from the "rocket-scalpel" to the "gene-meteorite" complex', *Social Science Information*, 37 (4), 731-51.
- (2012), 'Public Attention to Science 1820-2010 – a 'Longue Duree' Picture', in S. Rödder, M. Franzen, and P. Weingart (eds.), *The sciences' media connection – Public communication and its repercussions* (Springer Science), 35-57.
- Begley, C. G. and Ellis, L. M. (2012), 'Drug development: Raise standards for preclinical cancer research', *Nature*, 483 (7391), 531-3.
- Begley, C. G. and Ioannidis, J. P. (2015), 'Reproducibility in science: improving the standard for basic and preclinical research', *Circ Res*, 116 (1), 116-26.
- Bérard, Y. and Roger, A. (2015), 'Ronds-points théoriques et passages à niveaux analytiques. La sociologie politique peut-elle rencontrer la sociologie des sciences ?', *Politix*, 111 (3), 9-26.
- Bertamini, M. and Munafo, M. R. (2012), 'Bite-Size Science and Its Undesired Side Effects', *Perspect Psychol Sci*, 7 (1), 67-71.
- Besley, J. C. and Nisbet, M. C. (2013), 'How Scientists View the Public, the Media and the Political Process', *Public Underst Sci*, 22 (6), 644-59.
- Bissell, M. (2013), 'Reproducibility: The risks of the replication drive', *Nature*, 503 (7476), 333-4.
- Boltanski, L. and Maldidier, P. (1970), 'Carrière scientifique, morale scientifique et vulgarisation', *Inform. sci. soc.*, 9 (3), 99-118.

- Bonneuil, C. (2004), 'Les transformations des rapports entre sciences et société en France depuis la Seconde Guerre mondiale : un essai de synthèse', *Sciences, Médias et Société* (Cadrages historiques et comparatifs; Lyon: Actes), 15.
- Borup, M., et al. (2006), 'The sociology of expectations in Science and Technology', *Technology analysis & strategic management*, 18 (3/4), 285-98.
- Bourdaa, M., et al. (2015), 'Does television reflect the evolution of scientific knowledge? The case of attention deficit hyperactivity disorder coverage on French television', *Public Underst Sci*.
- Boutron, I., et al. (2010), 'Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes', *JAMA*, 303 (20), 2058-64.
- Brechman, J. M., Lee, C. J., and Cappella, J. N. (2009), 'Lost in Translation? A Comparison of Cancer-Genetics Reporting in the Press Release and its Subsequent Coverage in Lay Press', *Sci Commun*, 30 (4), 453-74.
- Brembs, B., Button, K., and Munafò, M. (2013), 'Deep impact: unintended consequences of journal rank', *Front Hum Neurosci*, 7, 291.
- Brischoux, F. and Angelier, F. (2015), 'Academia's never-ending selection for productivity', *Scientometrics*, 103 (1), 333-36.
- Brown, N. (2003), 'Hope against hype-accountability in biopasts, presents and futures', *Science studies*, 16 (2), 3-21.
- Brown, P. (2012), 'Nothing but the truth. Are the media as bad at communicating science as scientists fear?', *EMBO Rep*, 13 (11), 964-7.
- Bruno, I. (2012), 'À vos marques®, prêts cherchez ! La stratégie européenne de Lisbonne, vers un marché de la recherche.', *Mouvements*, 3 (71), 36-39.
- (2013), 'Ne cherchez plus, innovez!', *Revue Française de Socio-Economie*, 1 (11), 9-24.
- Brunt, M. E., et al. (2003), 'Mass media release of medical research results: an analysis of antihypertensive drug prescribing in the aftermath of the calcium channel blocker scare of March 1995', *J Gen Intern Med*, 18 (2), 84-94.
- Bubela, T., et al. (2009), 'Science communication reconsidered', *Nat Biotechnol*, 27 (6), 514-8.
- Bucchi, M. (2015), 'Norms, competition and visibility in contemporary science: the legacy of Robert K. Merton', *Journal of classical Sociology*, 15 (3), 233-52.
- Button, K. S. (2016), 'Statistical Rigor and the Perils of Chance', *eNeuro*, 3 (4).
- Button, K. S., et al. (2013), 'Power failure: why small sample size undermines the reliability of neuroscience', *Nat Rev Neurosci*, 14 (5), 365-76.
- Callahan, M., Wears, R. L., and Weber, E. (2002), 'Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals', *JAMA*, 287 (21), 2847-50.
- Callon, M. (1998), 'Des différentes formes de démocratie technique', *Annales des Mines*, 9, 63-73.
- Callon, M., Lascoumes, P., and Barthe, Y. (2014), *Agir dans un monde incertain. Essai sur la démocratie technique* ed. Essais (Points).
- Casadevall, A. and Fang, F. C. (2010), 'Reproducible science', *Infect Immun*, 78 (12), 4972-5.
- (2014b), 'Causes for the persistence of impact factor mania', *MBio*, 5 (2), e00064-14.
- Caulfield, T. (2004), 'Biotechnology and the popular press: hype and the selling of science', *Trends Biotechnol*, 22 (7), 337-9.
- Caulfield, T. and Condit, C. (2012), 'Science and the sources of hype', *Public Health Genomics*, 15 (3-4), 209-17.
- Caulfield, T., et al. (2016), 'SCIENTIFIC COMMUNITY. Confronting stem cell hype', *Science*, 352 (6287), 776-7.
- Chalmers, I., et al. (2014), 'How to increase value and reduce waste when research priorities are set', *Lancet*, 383 (9912), 156-65.

- Champagne, P. and Marchetti, D. (1994), 'l'information médicale sous contrainte. A propos du "scandale du sang contaminé"', *Actes de la recherche en sciences sociales*, 101-102, 40-62.
- Chapman, S., Nguyen, T. N., and White, C. (2007), 'Press-released papers are more downloaded and cited', *Tob Control*, 16 (1), 71.
- Chew, M., Villanueva, E. V., and Van Der Weyden, M. B. (2007), 'Life and times of the impact factor: retrospective analysis of trends for seven medical journals (1994-2005) and their Editors' views.', *J R Soc Med*, 100 (3), 142-50.
- Clarke, C. E. (2011), 'A case of conflicting norms? Mobilizing and accountability information in newspaper coverage of the autism-vaccine controversy', *Public Underst Sci*, 20 (5), 609-26.
- Collins, F. S. and Tabak, L. A. (2014), 'Policy: NIH plans to enhance reproducibility', *Nature*, 505 (7485), 612-3.
- Collins, H. (1987), 'Certainty and the Public Understanding of Science: Science on Television', *Social studies of Science*, 17 (4), 689-713.
- (2009), 'We cannot live by scepticism alone', *Nature*, 458 (7234), 30.
- Collins, H. M and Pinch, T. (2001), *Tout ce que vous devriez savoir sur la science* (Paris: Seuil).
- Collins, H. M. and Evans, R.J. (2002), 'The Third Wave of Science Studies: Studies of Expertise and Experience', *Social studies of Science*, 32 (2), 235-96.
- Collins, H.M. (1983), 'The Sociology of Scientific Knowledge: Studies of Contemporary Science', *Annual review of sociology*, 9, 265-85.
- Collins, H.M., Evans, R., and Weinel, M. (2017), 'STS as science or politics?', *Social Studies of science*, 47 (4), 580-86.
- Comby, J.B. (2015), *La question climatique. Genèse et dépolitisation d'un problème public* (Cours et travaux; Paris).
- Condit, C. (2004), 'Science reporting to the public: does the message get twisted?', *CMAJ*, 170 (9), 1415-6.
- Conner, C.S. (1986), 'Conflicting data in the literature: a true dilemma. ', *Annals of pharmacotherapy*, 20 (6), 444-45.
- Conrad, P. (1999), 'Uses of expertise: sources, quotes, and voice in the reporting of genetics in the news', *Public Underst Sci*, 8 (4), 285-302.
- Cooper, C. P. and Yukimura, D. (2002), 'Science writers' reactions to a medical "breakthrough" story', *Social science & Medicine*, 54 (12), 1887-96.
- Corbett, J. B. and Mori, M. (1999), 'Medicine, Media, and Celebrities: News Coverage of Breast Cancer, 1960-1995', *J&MC Quarterly*, 76 (2), 229-49.
- Couzin-Frankel, J. (2013), 'The power of negative thinking', *Science*, 342 (6154), 68-9.
- de Semir, V. (1996), 'What is newsworthy?', *Lancet*, 347 (9009), 1163-6.
- (2010), 'Science Communication & Science Journalism Meta-review ', in *Media for science* (ed.).
- de Semir, V., Ribas, C., and Revuelta, G. (1998), 'Press releases of science journal articles and subsequent newspaper stories on the same topic.', *JAMA*, 280 (3), 294-5.
- de Vries, R., Anderson, M. S., and Martinson, B. C. (2006), 'Normal Misbehavior: Scientists Talk about the Ethics of Research', *J Empir Res Hum Res Ethics*, 1 (1), 43-50.
- de Winter, J. C. and Dodou, D. (2015), 'A surge of p-values between 0.041 and 0.049 in recent decades (but negative results are increasing rapidly too)', *PeerJ*, 3, e733.
- Dentzer, S. (2009), 'Communicating medical news--pitfalls of health care journalism', *N Engl J Med*, 360 (1), 1-3.
- Dienes, Z. (2014), 'Using Bayes to get the most out of non-significant results', *Front Psychol*, 5, 781.
- DiMasi, J. A., et al. (2010), 'Trends in risks associated with new drug development: success rates for investigational drugs', *Clin Pharmacol Ther*, 87 (3), 272-7.

- Dornan, C. (1990), 'Some problems in conceptualizing the issue of "science and the media"', *Critical Studies in Mass Communication*, 7 (1), 48-71.
- Dubois, M. (2013), 'Introduction', *l'année sociologique*, 63 (2), 345-57.
- Dudo, A., Dunwoody, S., and Scheufele, D. A. (2011), 'The emergence of nano news: Tackling thematic trends and changes in U.S. newspaper coverage of nanotechnology.', *Journalism & Mass Communication Quarterly* 88 (1), 55-75.
- Dumas-Mallet, E., et al. (2017), 'Poor replication validity of biomedical association studies reported by newspapers', *PLoS One*, 12 (2), e0172650.
- (2017soumis), 'How newspapers deal with scientific uncertainty when reporting emergent biomedical findings', *Sci Commun*, Soumis.
- Dumas-Mallet, E., et al. (2016), 'Replication Validity of Initial Association Studies: A Comparison between Psychiatry, Neurology and Four Somatic Diseases', *PLoS One*, 11 (6), e0158064.
- Dunwoody, S. and Scott, B.T. (1982), 'Scientists as mass media sources', *Journalism quarterly*.
- Dunwoody, S. and Ryan, M. (1987), 'The Credible Scientific Source', *Journalism & Mass Communication Quarterly*, 64 (1), 21-27.
- Durant, J. R., Evans, G. A., and Thomas, G. P. (1989), 'The public understanding of science', *Nature*, 340 (6228), 11-4.
- Ebersole, C. R., Axt, J. R., and Nosek, B. A. (2016), 'Scientists' Reputations Are Based on Getting It Right, Not Being Right', *PLoS Biol*, 14 (5), e1002460.
- Edwards, M. A. and Roy, S. (2017), 'Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition', *Environ Eng Sci*, 34 (1), 51-61.
- Egger, M., et al. (1997), 'Bias in meta-analysis detected by a simple, graphical test', *BMJ*, 315 (7109), 629-34.
- Elzinga, A. (1997), 'The science-society contract in historical transformation: with special reference to "epistemic drift"', *Social Science Information*, 36 (3), 411-45.
- Entwistle, V. (1995), 'Reporting research in medical journals and newspapers', *BMJ*, 310 (6984), 920-3.
- Epstein, S. (1995), 'The Construction of Lay Expertise: AIDS Activism and the Forging of Credibility in the Reform of Clinical Trials', *Science, Technology & Human Values*, 20 (4), 408-37.
- Eurobaromètre (2007), 'La recherche scientifique dans les médias', (282).
- Eyck, T. A. T and Williment, M. (2003), 'The National Media and Things Genetic Coverage in the New York Times (1971-2001) and the Washington Post (1977-2001)', *Sci Commun*, 25 (2), 129-52.
- Fanelli, D. (2009), 'How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data', *PLoS One*, 4 (5), e5738.
- (2010), 'Do pressures to publish increase scientists' bias? An empirical support from US States Data', *PLoS One*, 5 (4), e10271.
- (2010b), '"Positive" results increase down the Hierarchy of the Sciences', *PLoS One*, 5 (4), e10068.
- (2012), 'Negative results are disappearing from most disciplines and countries', *Scientometrics*, 90 (3), 891-904.
- Fanelli, D., Costas, R., and Ioannidis, J. P. (2017), 'Meta-assessment of bias in science', *Proc Natl Acad Sci U S A*, 114 (14), 3714-19.
- Fischhoff, B. and Davis, A. L. (2014), 'Communicating scientific uncertainty', *Proc Natl Acad Sci U S A*, 111 Suppl 4, 13664-71.
- Flaysakier, J, D. (1997), 'Santé publique et responsabilité des médias', *Hermès*, 21, 135-43.
- Fouquet, A. and Perriault, J. (2010), 'Évaluation, politiques publiques, politique de recherche', *Communication et Organisation*, 38, 29-40.

- Franzen, M. (2012), 'Making science news: The press relations of scientific journals and implications for scholarly communication.', in S. Rödder, M. Franzen, and P. Weingart (eds.), *The sciences' media connection – Public communication and its repercussions* (Springer Science), 333-52.
- Fraser, V. J. and Martin, J. G. (2009), 'Marketing data: has the rise of impact factor led to the fall of objective language in the scientific article?', *Respir Res*, 10, 35.
- Freedman, L. P., Cockburn, I. M., and Simcoe, T. S. (2015), 'The Economics of Reproducibility in Preclinical Research', *PLoS Biol*, 13 (6), e1002165.
- Friedman, S.M. (2011), 'Three Mile Island, Chernobyl, and Fukushima: An analysis of traditional and new media coverage of nuclear accidents and radiation', *Bulletin of the atomic scientists*, 67 (5), 55-65.
- Friedman, S.M., Dunwoody, S., and Rogers, C.L. (1986), *Scientists and Journalists Reporting science as news* (New-york: Free Press (Macmillian)).
- (1999), *Communicating Uncertainty: media coverage of new and Controversial science* (USA: Lawrence Erlbaum Associates).
- Galtung, J. and Ruge, M.H. (1965), 'The structure of foreign news. The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers', *Journal of Peace Research*, 2 (1), 64-90.
- Gan, H. K., et al. (2012), 'Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer', *J Natl Cancer Inst*, 104 (8), 590-8.
- Gascoigne, T and Metcalfe, J. (1997), 'Incentives and Impediments to Scientists Communicating Through the Media', *Sci Commun*, 18 (3), 265-82.
- Geller, G., et al. (2005), 'Scientists' and science writers' experiences reporting genetic discoveries: toward an ethic of trust in science journalism', *Genet Med*, 7 (3), 198-205.
- Gibbons, M. (1999), 'Science's new social contract with society', *Nature*, 402 (6761 Suppl), C81-4.
- Gibbons, M., et al. (1994), 'The new production of knowledge : the dynamics of science and research in contemporary societies', *SAGE Publishing*.
- Glasziou, P., et al. (2014), 'Reducing waste from incomplete or unusable reports of biomedical research', *Lancet*, 383 (9913), 267-76.
- Godin, B. and Gingras, Y. (2002), 'The experimenters' regress: from skepticism to argumentation', *Studies in History and Philosophy of science*, 33 (1), 133-48.
- Gonon, F. (2009), 'The dopaminergic hypothesis of attention-deficit/hyperactivity disorder needs re-examining', *Trends Neurosci*, 32 (1), 2-8.
- Gonon, F., Bezaud, E., and Boraud, T. (2011), 'Misrepresentation of neuroscience data might give rise to misleading conclusions in the media: the case of attention deficit hyperactivity disorder', *PLoS One*, 6 (1), e14618.
- Gonon, F., Konsman, J. P., and Boraud, T. (2014), 'Neurosciences et médiatisation : entre argumentation de la preuve et rhétorique de la promesse', in B. Chamaç and B. Moutaud (eds.), *Neurosciences et société - Enjeux des savoirs et pratiques sur le cerveau* (Armand Colin/r Recherches).
- Gonon, F., et al. (2012), 'Why most biomedical findings echoed by newspapers turn out to be false: the case of attention deficit hyperactivity disorder', *PLoS One*, 7 (9), e44275.
- Greenberg, S. A. (2009), 'How citation distortions create unfounded authority: analysis of a citation network', *BMJ*, 339, b2680.
- Grilli, R., Ramsay, C., and Minozzi, S. (2002), 'Mass media interventions: effects on health services utilisation', *Cochrane Database Syst Rev*, (1), CD000389.
- Guenther, L. and Ruhrmann, G. (2013), 'Science journalists' selection criteria and depiction of nanotechnology in German media', *Journal of Science communication*, 12 (03), 1-17.
- (2016), 'Scientific evidence and mass media: Investigating the journalistic intention to represent scientific uncertainty', *Public Underst Sci*, 25 (8), 927-43.

- Haneef, R., et al. (2015), 'Interpretation of Results of Studies Evaluating an Intervention Highlighted in Google Health News: A Cross-Sectional Study of News', *PLoS One*, 10 (10), e0140889.
- Haneef, R., et al. (2017), 'Factors associated with online media attention to research: a cohort study of articles evaluating cancer treatments', *Research Integrity and Peer Review*, 2 (1), 9.
- Hanna, M. (2015), 'Matching taxpayer funding to population health needs', *Circ Res*, 116 (8), 1296-300.
- Hansen, A. (1994), 'Journalistic practices and science reporting in the British press', *Public Underst Sci*, 3 (2), 111-34.
- (2016), 'The changing uses of accuracy in science communication', *Public Underst Sci*, 25 (7), 760-74.
- Haran, J. and Kitzinger, J. (2009), 'Modest witnessing and managing the boundaries between science and the media: A case study of breakthrough and scandal', *Public Underst Sci*, 18 (6), 634-52.
- Harris, R. (2017), *Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions* (Basic Books).
- Hartz, J. and Chappell, R. (1997), *Worlds Apart: How the Distance Between Science and Journalism Threatens America's Future* (Nashville: First amendment center).
- Hessels, L.K, van Lente, H. (2008), 'Re-thinking new knowledge production: A literature review and a research agenda', *Research policy*, 37 (4), 740-60.
- Hilgartner (1990), 'The Dominant View of Popularization: Conceptual Problems, Political Uses', *Social Studies of science*, 20 (3), 519-39.
- Hirschhorn, J. N., et al. (2002), 'A comprehensive review of genetic association studies', *Genet Med*, 4 (2), 45-61.
- Hodgetts, D., et al. (2008), 'Constructing health news: possibilities for a civic-oriented journalism', *Health (London)*, 12 (1), 43-66.
- Holtzman, N. A., et al. (2005), 'The quality of media reports on discoveries related to human genetic diseases', *Community Genet*, 8 (3), 133-44.
- Hornmoen, H. (2010), 'What researchers now can tell us : Representing scientific uncertainty in journalism', *Observatorio*, 3 (4), 74-100.
- Horton, R. (2015), 'Offline: What is medicine's 5 sigma?', *Lancet*, 385 (9976), 1380.
- Iaboli, L., et al. (2010), 'The unbearable lightness of health science reporting: a week examining Italian print media', *PLoS One*, 5 (3), e9829.
- Ioannidis, J. P. (2005), 'Why most published research findings are false', *PLoS Med*, 2 (8), e124.
- (2005b), 'Contradicted and initially stronger effects in highly cited clinical research', *JAMA*, 294 (2), 218-28.
- (2008), 'Why most discovered true associations are inflated', *Epidemiology*, 19 (5), 640-8.
- (2008b), 'Effect of formal statistical significance on the credibility of observational associations', *Am J Epidemiol*, 168 (4), 374-83; discussion 84-90.
- (2011), 'More time for research: fund people not projects', *Nature*, 477 (7366), 529-31.
- (2011b), 'Excess significance bias in the literature on brain volume abnormalities', *Arch Gen Psychiatry*, 68 (8), 773-80.
- (2011d), 'Commentary: Adjusting for bias: a user's guide to performing plastic surgery on meta-analyses of observational studies', *Int J Epidemiol*, 40 (3), 777-9.
- (2013), 'This I believe in genetics: discovery can be a nuisance, replication is science, implementation matters', *Front Genet*, 4, 33.
- (2014c), 'How to make more published research true', *PLoS Med*, 11 (10), e1001747.
- Ioannidis, J. P. and Panagiotou, O. A. (2011c), 'Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses', *JAMA*, 305 (21), 2200-10.

- Ioannidis, J. P., et al. (2001b), 'Replication validity of genetic association studies', *Nat Genet*, 29 (3), 306-9.
- Ioannidis, J. P., et al. (2003), 'Genetic associations in large versus small studies: an empirical assessment', *Lancet*, 361 (9357), 567-71.
- Ioannidis, J. P., et al. (2014), 'Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention', *Trends Cogn Sci*, 18 (5), 235-41.
- Ioannidis, J. P., et al. (2014b), 'Increasing value and reducing waste in research design, conduct, and analysis', *Lancet*, 383 (9912), 166-75.
- Ioannidis, J. P., et al. (2009), 'Repeatability of published microarray gene expression analyses', *Nat Genet*, 41 (2), 149-55.
- Ioannidis, J.P. and Lau, J. (2001), 'Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses', *Proc Natl Acad Sci U S A*, 98 (3), 831-6.
- Iqbal, S. A., et al. (2016), 'Reproducible Research Practices and Transparency across the Biomedical Literature', *PLoS Biol*, 14 (1), e1002333.
- Irwin, A. (2006), 'The Politics of Talk: Coming to Terms with the 'New' Scientific Governance', *Social Studies of science*, 36 (2), 299-320.
- Jannot, A. S., et al. (2013), 'Citation bias favoring statistically significant studies was present in medical research', *J Clin Epidemiol*, 66 (3), 296-301.
- Jasanoff, S. (1996), 'Beyond Epistemology: Relativism and Engagement in the Politics of Science', *Social Studies of science*, 26 (2), 393-418.
- Jensen, E. (2010), 'Between credulity and scepticism: envisaging the fourth estate in 21st-century science journalism', *Media, culture & society*, 32 (4), 615-30.
- Jensen, J. D. and Hurley, R. J. (2012), 'Conflicting stories about public scientific controversies: Effects of news convergence and divergence on scientists' credibility', *Public Underst Sci*, 21 (6), 689-704.
- Jensen, J. D., et al. (2013), 'Against conventional wisdom: when the public, the media, and medical practice collide', *BMC Medical Informatics and Decision Making*, 13(suppl3) (S4).
- Jensen, P., et al. (2008), 'Scientists who engage with society perform better academically', *Science and Public Policy*, 35 (7), 527-41.
- Jinek, M., et al. (2012), 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*, 337 (6096), 816-21.
- Johnson, V. E. (2013), 'Revised standards for statistical evidence', *Proc Natl Acad Sci U S A*, 110 (48), 19313-7.
- Jouvenet, M. (2011), 'Profession scientifique et instruments politiques : l'impact du financement « sur projet » dans des laboratoires de nanosciences', *Sociologie du travail*, 53 (2), 234-52.
- Karmarkar, U.R. and Tormala, Z.L. (2010), 'Believe Me, I Have No Idea What I'm Talking About: The Effects of Source Certainty on Consumer Involvement and Persuasion', *Journal of Consumer Research*, 36 (6), 1033-49.
- Kass, R. E., et al. (2016), 'Ten Simple Rules for Effective Statistical Practice', *PLoS Comput Biol*, 12 (6), e1004961.
- Kerr, N. L. (1998), 'HARKing: hypothesizing after the results are known', *Pers Soc Psychol Rev*, 2 (3), 196-217.
- Kitzinger, Jenny and Reilly, Jacquie (1997), 'The rise and fall of risk reporting: Media coverage of human genetics research, 'false memory syndrome' and 'mad cow disease'.', *European journal of communication*, 12 (3), 319-50.
- Kivimaki, M., et al. (2014), 'Don't let the truth get in the way of a good story: an illustration of citation bias in epidemiologic research', *Am J Epidemiol*, 180 (4), 446-8.
- Kua, E., Reder, M., and Grossel, M. J. (2004), 'Science in the news: a study of reporting genomics', *Public Underst Sci*, 13 (3), 309-22.
- Kulkarni, A. V., Busse, J. W., and Shams, I. (2007), 'Characteristics associated with citation rate of the medical literature', *PLoS One*, 2 (5), e403.

- Lam, A. (2010), 'From 'Ivory Tower Traditionalists' to 'Entrepreneurial Scientists'? Academic Scientists in Fuzzy University-Industry Boundaries', *Social studies of Science*, 40 (2), 307-40.
- Larivière, V., Lozano, G.A., and Gingras, Y. (2014), 'Are elite journals declining?', *Journal of the Association for Information Science and Technology*, 65 (4), 649-55.
- Laurance, J. (1998), 'This is what the game is about', *Lancet*, 351, 1727-28.
- Lawrence, P. A. (2003), 'The politics of publication', *Nature*, 422 (6929), 259-61.
- (2007), 'The mismeasurement of science', *Curr Biol*, 17 (15), R583-5.
- (2009), 'Real lives and white lies in the funding of scientific research: the granting system turns young scientists into bureaucrats and then betrays them', *PLoS Biol*, 7 (9), e1000197.
- Lehmkuhl, M. and Peters, H. P. (2016), 'Constructing (un-)certainty: An exploration of journalistic decision-making in the reporting of neuroscience', *Public Underst Sci*, 25 (8), 909-26.
- Lehrer, J. (2010), 'The truth wears off: is there something wrong with the scientific method?', *The New Yorker*, p. 52-57.
- Lettre de l'Institut National des Sciences Biologiques (2016), (CNRS-INSB).
- Lewandowsky, S. and Bishop, D. (2016), 'Research integrity: Don't let transparency damage science', *Nature*, 529 (7587), 459-61.
- Lewis, J., Williams, A., and Franklin, B. (2008), 'A compromised fourth estate? UK news journalism, public relations and news sources', *Journalism Studies*, 9 (1), 1-20.
- Logan, R. A., Zengjun, P., and Wilson, N, F. (2000), 'Science and Medical Coverage in the Los Angeles Times and The Washington Post: A Six-Year Perspective', *Sci Commun*, 22 (1), 5-26.
- Macleod, M. R., et al. (2014), 'Biomedical research: increasing value, reducing waste', *Lancet*, 383 (9912), 101-4.
- Maille, ME., Saint-Charles, J., and Lucotte, M. (2010), 'The gap between scientists and journalists: the case of mercury science in Québec's press', *Public Underst Sci*, 19 (1), 70-79.
- Marchetti, D. (2002), 'Les sous champs spécialisés du journalisme', *Réseaux*, 1 (111), 22-55.
- (2010), *Quand la santé devient médiatique. Les logiques de production de l'information dans la presse* (Grenoble: Presses Universitaires de Grenoble).
- Marshall, E. (1998), 'The power of the front page of The New York Times', *Science*, 280 (5366), 996-7.
- Martinson, B. C., Anderson, M. S., and de Vries, R. (2005), 'Scientists behaving badly', *Nature*, 435 (7043), 737-8.
- Master, Z. and Resnik, D. B. (2013), 'Hype and public trust in science', *Sci Eng Ethics*, 19 (2), 321-35.
- Merton, R. K. (1973), 'The normative structure of Science', *The sociology of science: theoretical and empirical investigations* (Chicago: University of Chicago press), 267-78.
- Ministère de l'Education Nationale, de l'Enseignement supérieur et de la recherche (2015), 'Stratégie nationale de recherche France Europe 2020', (Paris).
- Mitroff, I. I. (1976), 'Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists', *American sociology review*, 39 (4), 579-95.
- Mobley, A., et al. (2013), 'A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic', *PLoS One*, 8 (5), e63221.
- Moreira, T. and Palladino, P. (2005), 'Between truth and hope : on parkinson's disease, neurotransplantation and the production of the "self"', *History of the human sciences*, 18 (3), 55-82.

- Mountcastle-Shah, E., et al. (2003), 'Assessing mass media reporting of disease-related genetic discoveries', *Sci Commun*, 24 (4), 458-78.
- Moynihan, R., et al. (2000), 'Coverage by the news media of the benefits and risks of medications', *N Engl J Med*, 342 (22), 1645-50.
- Munafò, M., et al. (2017), 'A manifesto for reproducible science', *Nature Human Behaviour*, 1, 0021.
- Munafò, M., et al. (2014), 'Scientific rigor and the art of motorcycle maintenance', *Nat Biotechnol*, 32 (9), 871-3.
- Munafò, M. R., Stothart, G., and Flint, J. (2009), 'Bias in genetic association studies and impact factor', *Mol Psychiatry*, 14 (2), 119-20.
- National Academies of Sciences, Engineering and Medicine (2017), *Communicating Science Effectively: A Research Agenda* (Washington: National Academies press).
- Nelkin, D. (1995), *Selling science: How the press covers science and technology* (New York: W.H. Freeman and Company).
- (1996), 'An uneasy relationship: the tensions between medicine and the media', *Lancet*, 347 (9015), 1600-3.
- Neveu, E. (2009), *Sociologie du Journalisme*, ed. Repères (La découverte).
- Nisbet, M. C. and Lewenstein, B. V. (2002b), 'Biotechnology and the American Media. The Policy Process and the Elite Press, 1970 to 1999', *Sci Commun*, 23 (4), 359-91.
- Nosek, B. A., Spies, J. R., and Motyl, M. (2012), 'Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability', *Perspect Psychol Sci*, 7 (6), 615-31.
- Nowotny, H. (2004), 'The Potential of Transdisciplinarity', in CNRS (ed.), *Rethinking interdisciplinarity*.
- Nowotny, H., Scott, P., and Gibbons, M. (2003), 'Introduction: Mode 2' Revisited: The New Production of Knowledge', *Minerva*, 41 (3), 179-94.
- Nuzzo, R. (2014), 'Scientific method: statistical errors', *Nature*, 506 (7487), 150-2.
- Open Science Collaboration (2015), 'PSYCHOLOGY. Estimating the reproducibility of psychological science', *Science*, 349 (6251), aac4716.
- Oransky, I. (2015), 'How Publish or Perish Promotes Inaccuracy in Science -- and Journalism', *AMA J Ethics*, 17 (12), 1172-5.
- Osterloh, M., Frey, B. S., and Homberg, F. (2008), 'Le chercheur et l'obligation de rendre des comptes', *Gérer et comprendre*, 91 (1), 48-54.
- Pellechia, M. G. (1997), 'Trends in science coverage: a content analysis of three US newspapers', *Public Underst Sci*, 6 (1), 49-68.
- Pellizzoni, L. (2003), 'Knowledge, Uncertainty and the Transformation of the Public Sphere', *European Journal of Social Theory*, 6 (3), 327-55.
- Pereira, T. V. and Ioannidis, J. P. (2011), 'Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects', *J Clin Epidemiol*, 64 (10), 1060-9.
- Pereira, T. V., Horwitz, R. I., and Ioannidis, J. P. (2012), 'Empirical evaluation of very large treatment effects of medical interventions', *JAMA*, 308 (16), 1676-84.
- Pestre, D. (2004), 'Thirty Years of Science Studies: Knowledge, Society and the Political', *History and Technology*, 20 (4), 351-69.
- Peters, H. P. (1995), 'The interaction of journalists and scientific experts: co-operation and conflict between two professional cultures', *Media, Culture & Society*, 17, 31-48.
- (2012), 'Le journalisme scientifique : « médier » la relation entre science et société', *Cahiers de journalisme*, 24, 14-31.
- (2012b), *Scientific sources and the mass media: forms and consequences of medialization*, ed. The sciences' media connection- Public communication and its repercussions (Sociology of sciences yearbook, 28).
- (2013), 'Gap between science and media revisited: scientists as public communicators', *Proc Natl Acad Sci U S A*, 110 Suppl 3, 14102-9.

- Peters, H. P., et al. (2008), 'Science communication. Interactions with the mass media', *Science*, 321 (5886), 204-5.
- Peters, H. P., et al. (2008b), 'Science-Media Interface : it's Time to Reconsider', *Sci Commun*, 30 (2), 266-76.
- Petersen, I., Heinrichs, H., and Peters, H.P. (2010), 'Mass-Mediated Expertise as Informal Policy Advice', *Science, Technology, & Human Values*, 35 (6), 865-87.
- Phillips, D. P., et al. (1991), 'Importance of the lay press in the transmission of medical knowledge to the scientific community', *N Engl J Med*, 325 (16), 1180-3.
- Pinch, T. (1981), 'The Sun-Set: the presentation of certainty on scientific life', *Social Studies of science*, 11 (1), 131-58.
- Price, D.J. De S. (1986), *Little Science, Big Science.... and Beyond* (New York: Columbia University Press).
- Priest, S, H. (2001), 'Misplaced Faith', *Sci Commun*, 23 (2), 97-110.
- Prinz, F., Schlange, T., and Asadullah, K. (2011), 'Believe it or not: how much can we rely on published data on potential drug targets?', *Nat Rev Drug Discov*, 10 (9), 712.
- Rafols, I., Porter, A, L., and Leydesdorff, L. (2010), 'Science overlay maps: A new tool for research policy and library management.', *Journal of the american society for information science ans technology*, 61 (9), 1871-87.
- Reed, R. (2001), '(Un-)Professional discourse? Journalists' and scientists' stories about science in the media', *Journalism*, 2 (3), 279-98.
- Reich, E. S. (2013), 'Science publishing: The golden club', *Nature*, 502 (7471), 291-3.
- Rodder, S. (2009), 'Reassessing the concept of a medialization of science: a story from the "book of life"', *Public Underst Sci*, 18 (4), 452-63.
- Ruhrmann, G., et al. (2015), 'Frames of scientific evidence: How journalists represent the (un)certainty of molecular medicine in science television programs', *Public Underst Sci*, 24 (6), 681-96.
- Rusconi, E. and Mitcher-Nissen, T. (2014), 'The role of expectations, hype and ethics in neuroimaging and neuromodulation futures', *Frontiers in systems neuroscience*, 8.
- Russell, C. (1999), 'Living can be hazardous to your health: how the news media cover cancer risks', *J Natl Cancer Inst Monogr*, (25), 167-70.
- Samson, A (2005), 'Plagiarism and fabulism: Dishonesty in the newsroom', *Pacific Journalism Review*, 11 (2), 84-100.
- Samuel, G. and Williams, C. (2017), 'UK science press officers, professional vision and the generation of expectations', *Public Underst Sci*, 26 (1), 55-69.
- Sarewitz, D. (2012), 'Beware the creeping cracks of bias', *Nature*, 485 (7397), 149.
- Schafer, M. S. (2009), 'From Public Understanding to Public Engagement: An Empirical Assessment of Changes in Science Coverage', *Sci Commun*, 30 (4), 475-505.
- (2011b), 'Sources, characteristics and effects of mass media communication on science: A review of the literature, current trends and areas for future Research', *Sociology compass*, 5/6, 399-412.
- Schafer, Maximilian. (2011), 'Science journalism and fact checking', *Journal of Science communication*, 10 (4).
- Schekman, R. (2013), 'How journals like Nature, Cell and Science are damaging science', *The Guardian*.
- Schneider, J. (2010), 'Making space for the "Nuances of Truth": Communication and Uncertainty at an Environmental Journalists' Workshop', *Sci Commun*, 32 (2), 171-201.
- Schooler, J. (2011), 'Unpublished results hide the decline effect', *Nature*, 470 (7335), 437.
- Schwartz, L. M. and Woloshin, S. (2004), 'The media matter: a call for straightforward medical reporting', *Ann Intern Med*, 140 (3), 226-8.
- Schwartz, L. M., et al. (2012), 'Influence of medical journal press releases on the quality of associated newspaper coverage: retrospective cohort study', *BMJ*, 344, d8164.

- Schwitzer, G. (2008), 'How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories', *PLoS Med*, 5 (5), e95.
- Sessions, L. (2003), 'Verification and balance in science news: how the New Zealand mass media report scientific claims', (University of Canterbury NZ).
- Sharma, V., et al. (2003), 'Influence of the news media on diagnostic testing in the emergency department', *Arch Pediatr Adolesc Med*, 157 (3), 257-60.
- Shinn, T. (2002), 'Nouvelle Production du Savoir et Triple Hélice [Tendances du prêt-à-penser les sciences]', *Actes de la recherche en sciences sociales*, 141 (1), 21-30.
- Shoemaker, P. J. and Reese, S. D. (1996), *Mediating the message: theories of influences on mass media content* (2nd edition edn.; N.Y. 10606: Longan Publishers USA).
- Shoemaker, P. J., et al. (2001), 'Individual and routine forces in Gatekeeping', *J&MC Quarterly*, 78 (2), 233-46.
- Sicard, MN. (1997), 'Pratiques journalistiques et enjeux de la communication scientifique et technique', *Hermès*, 21, 149-55.
- Siebert, S., Machesky, L. M., and Insall, R. H. (2015), 'Overflow in science and its implications for trust', *Elife*, 4.
- Siegfried, T. (1991), 'Newspapers', in B Lewenstein, V. (ed.), *When science meets the public: proceedings of a workshop organized by the American Association for the Advancement of Science Committee on Public Understanding of Science and Technology*, (Washington: Committee on Public Understanding of Science and Technology), 113- 24.
- Sim, J. and Wright, C. C. (2005), 'The kappa statistic in reliability studies: use, interpretation, and sample size requirements', *Phys Ther*, 85 (3), 257-68.
- Simons, K. (2008), 'The misused impact factor', *Science*, 322 (5899), 165.
- Singer, E. (1990), 'A Question of Accuracy: How Journalists and Scientists Report Research on Hazards', *Journal of communication*, 40 (4), 102-16.
- Smaldino, P. E. and McElreath, R. (2016), 'The natural selection of bad science', *R Soc Open Sci*, 3 (9), 160384.
- Smith, R. (2006), 'Medical journals and the mass media: moving from love and hate to love', *J R Soc Med*, 99 (7), 347-52.
- Steneck, N. H. (2006), 'Fostering integrity in research: definitions, current knowledge, and future directions', *Sci Eng Ethics*, 12 (1), 53-74.
- Sterne, J. A., Egger, M., and Smith, G. D. (2001), 'Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis', *BMJ*, 323 (7304), 101-5.
- Stocking, S. H. and Holstein, L. W. (2009), 'Manufacturing doubt: journalists' roles and the construction of ignorance in a scientific controversy', *Public Underst Sci*, 18 (1), 23-42.
- Stryker, J. E. (2002), 'Reporting medical information: effects of press releases and newsworthiness on medical journal articles' visibility in the news media', *Prev Med*, 35 (5), 519-30.
- Suleski, J. and Ibaraki, M (2010), 'Scientists are talking, but mostly to each other: a quantitative analysis of research represented in mass media', *Public Underst Sci*, 19 (1), 1115-25.
- Sumner, P., et al. (2016), 'Exaggerations and Caveats in Press Releases and Health-Related Science News', *PLoS One*, 11 (12), e0168217.
- Sumner, P., et al. (2014), 'The association between exaggeration in health related science news and academic press releases: retrospective observational study', *BMJ*, 349, g7015.
- Tajika, A., et al. (2015), 'Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up', *Br J Psychiatry*, 207 (4), 357-62.
- Tatsioni, A., Bonitsis, N. G., and Ioannidis, J. P. (2007), 'Persistence of contradicted claims in the literature', *JAMA*, 298 (21), 2517-26.

- The Economist (2012), 'Reporting science Journalistic deficit disorder - What newspapers don't say matters as much as what they do', *The Economist*, Sep 22.
- (2013a), 'Problems with scientific research. How science goes wrong?', *The Economist*.
- (2013b), 'Unreliable research Trouble at the lab', *The Economist*, Oct 18.
- The Royal Society (2006), 'Science communication excellence in science', (The royal society).
- Thompson, S., et al. (2011), 'A proposed method of bias adjustment for meta-analyses of published observational studies', *Int J Epidemiol*, 40 (3), 765-77.
- Tijdkink, J. K., et al. (2016), 'How do scientists perceive the current publication culture? A qualitative focus group interview study among Dutch biomedical researchers', *BMJ Open*, 6 (2), e008681.
- Treise, Debbie and Weigold, Michael F. (2002), 'Advancing Science Communication. A survey of science communicators', *Sci Commun*, 23 (3), 310-22.
- Trench, B. (2008), 'Towards an Analytical Framework of Science Communication Models', in D. Cheng, et al. (eds.), *Communicating science in social contexts: new models, new practices* (Springer Netherlands), 119-38.
- Trikalinos, T. A., et al. (2004b), 'Establishment of genetic associations for complex diseases is independent of early study findings', *Eur J Hum Genet*, 12 (9), 762-9.
- Trikalinos, T. A., et al. (2004), 'Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time', *J Clin Epidemiol*, 57 (11), 1124-30.
- Tristani-Potteaux, F. (1997), *Les journalistes scientifiques, médiateurs des savoirs*. (Paris: Economica).
- (2001), 'Du laboratoire au citoyen : les trois étapes de la communication scientifique', *CNRS info 394* (Special 20 ans), 7-8.
- Valenti, J. M. (1999), 'Commentary : how well do the scientists communicate to media? ', *Sci Commun*, 21 (2), 172-78.
- van Lente, H., Spitters, C., and Peine, A (2013), 'Comparing technological hype cycles: Towards a theory', *Technological Forecasting and Social Change*, 80 (8), 1615-28.
- van Trigt, A. M., et al. (1994), 'Journalists and their sources of ideas and information on medicines', *Soc Sci Med*, 38 (4), 637-43.
- Veneu, F., Amorim, L. H., and Massarani, L. (2008), 'Science journalism in Latin America: how the scientific information from a scientific source is accommodated when it is transformed into a journalistic story', *Journal of Science communication*, 7 (1).
- Vinkers, C. H., Tijdkink, J. K., and Otte, W. M. (2015), 'Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis', *BMJ*, 351, h6467.
- Wallace, B. C., et al. (2009), 'Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data', *BMC Med Res Methodol*, 9, 80.
- Wasserstein, R. L. and Lazar, N. A. (2016), 'The ASA's Statement on p-Values: Context, Process, and Purpose', *The American Statistician*, 70 (2), 129-33.
- Watts, S. (2014), 'Society needs more than wonder to respect science', *Nature*, 508 (7495), 151.
- Weigold, Michael F. (2001), 'Communicating Science : A Review of the Literature', *Sci Commun*, 23 (2), 164-93.
- Weingart, P. (1997), 'From "Finalization" to "Mode 2": old wine in new bottles?', *Social Science Information*, 36 (4), 591-613.
- Weingart, P. and Guenther, L. (2016), 'Science communication and the issue of trust', *Journal of Science communication*, 15 (05), C01.
- Weissgerber, T. L., et al. (2015), 'Beyond bar and line graphs: time for a new data presentation paradigm', *PLoS Biol*, 13 (4), e1002128.
- Weitkamp, E (2003), 'British newspapers privilege health and medicine topics over other science news', *Public Relations Review*, 29 (3), 321-33.

- (2014), 'On the roles of scientists, press officers and journalists', *Journal of Science communication*, 13 (03).
- White, D. M. (1950), 'The "gatekeeper": A case study in the selection of news.', *Journalism & Mass Communication Quarterly*, 27 (4), 383-90.
- Wilkie, T. (1996), 'Sources in science: who can we trust?', *Lancet*, 347 (9011), 1308-11.
- Wilkins, J. and Patterson, P. (1987), 'Risk Analysis and the Construction of News', *Journal of communication*, 37 (3), 80-92.
- Withey, S. B. (1959), 'Public Opinion about Science and Scientists', *Public Opinion Quarterly*, 23 (3), 382-88.
- Woloshin, S. and Schwartz, L. M. (2002), 'Press releases: translating research into news', *JAMA*, 287 (21), 2856-8.
- Woloshin, S., Schwartz, L. M., and Kramer, B. S. (2009), 'Promoting healthy skepticism in the news: helping journalists get it right', *J Natl Cancer Inst*, 101 (23), 1596-9.
- Woloshin, S., et al. (2009 b), 'Press releases by academic medical centers: not so academic?', *Ann Intern Med*, 150 (9), 613-8.
- Wolton, D. (1997), 'De la vulgarisation à la communication', *Hermès*, 21, 9-14.
- Wynne, B. (1996), 'Misunderstood misunderstandings: social identities and public uptake of science', in A. Irwin and B. Wynne (eds.), *Misunderstanding science? The public reconstruction of science and technology* (Cambridge University press).
- Yavchitz, A., et al. (2012), 'Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study', *PLoS Med*, 9 (9), e1001308.
- Yong, E. (2012), 'Replication studies: Bad copy', *Nature*, 485 (7398), 298-300.
- Young, N. S., Ioannidis, J. P., and Al-Ubaydli, O. (2008), 'Why current publication practices may distort science', *PLoS Med*, 5 (10), e201.
- Ziman, J. (1969), 'Information, Communication, Knowledge', *Nature*, 224, 318-24.
- (1991), 'Not knowing, needing to know and wanting to know', in B Lewenstein, V. (ed.), *When science meets the public: proceedings of a workshop organized by the American Association for the Advancement of Science Committee on Public Understanding of Science and Technology*, (Washington: Committee on Public Understanding of Science and Technology), 13-20.
- (1996), "'Post-Academic Science" : Constructing Knowledge with Networks and norms', *Science studies*, 9 (1), 67-80.
- (1996b), 'Is science losing its objectivity?', *Nature*, 382, 751-54.

ANNEXES

Annexe 1

RESEARCH ARTICLE

Replication Validity of Initial Association Studies: A Comparison between Psychiatry, Neurology and Four Somatic Diseases

Estelle Dumas-Mallet^{1,2,3}, Katherine Button⁴, Thomas Boraud^{1,2,6}, Marcus Munafo⁵, François Gonon^{1,2*}

1 CNRS, UMR 5293, Institute of Neurodegenerative diseases, Bordeaux, France, **2** University of Bordeaux, UMR 5293, Institute of Neurodegenerative diseases, Bordeaux, France, **3** CNRS, UMR 5116, Centre Emile Durkheim, Bordeaux, France, **4** University of Bath, Department of Psychology, Bath, United Kingdom, **5** MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom, **6** CHU Bordeaux, Bordeaux, France

* francois.gonon@u-bordeaux.fr



OPEN ACCESS

Citation: Dumas-Mallet E, Button K, Boraud T, Munafo M, Gonon F (2016) Replication Validity of Initial Association Studies: A Comparison between Psychiatry, Neurology and Four Somatic Diseases. PLoS ONE 11(6): e0158064. doi:10.1371/journal.pone.0158064

Editor: Ulrich S Tran, University of Vienna, School of Psychology, AUSTRIA

Received: March 17, 2016

Accepted: June 9, 2016

Published: June 23, 2016

Copyright: © 2016 Dumas-Mallet et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Context

There are growing concerns about effect size inflation and replication validity of association studies, but few observational investigations have explored the extent of these problems.

Objective

Using meta-analyses to measure the reliability of initial studies and explore whether this varies across biomedical domains and study types (cognitive/behavioral, brain imaging, genetic and “others”).

Methods

We analyzed 663 meta-analyses describing associations between markers or risk factors and 12 pathologies within three biomedical domains (psychiatry, neurology and four somatic diseases). We collected the effect size, sample size, publication year and Impact Factor of initial studies, largest studies (i.e., with the largest sample size) and the corresponding meta-analyses. Initial studies were considered as replicated if they were in nominal agreement with meta-analyses and if their effect size inflation was below 100%.

Results

Nominal agreement between initial studies and meta-analyses regarding the presence of a significant effect was not better than chance in psychiatry, whereas it was somewhat better in neurology and somatic diseases. Whereas effect sizes reported by largest studies and meta-analyses were similar, most of those reported by initial studies were inflated. Among the 256 initial studies reporting a significant effect ($p < 0.05$) and paired with significant meta-analyses, 97 effect sizes were inflated by more than 100%. Nominal agreement and effect size inflation varied with the biomedical domain and study type. Indeed, the replication rate

of initial studies reporting a significant effect ranged from 6.3% for genetic studies in psychiatry to 86.4% for cognitive/behavioral studies. Comparison between eight subgroups shows that replication rate decreases with sample size and “true” effect size. We observed no evidence of association between replication rate and publication year or Impact Factor.

Conclusion

The differences in reliability between biological psychiatry, neurology and somatic diseases suggest that there is room for improvement, at least in some subdomains.

Introduction

Many opinion and review articles have lamented the poor reproducibility of biomedical studies (e.g., [1–4]). However, relatively few empirical studies have quantitatively explored this issue across preclinical, clinical and association studies. The reproducibility of preclinical studies has been recently reviewed [5]; five studies estimated the prevalence of irreproducible results in preclinical studies to be between 51 and 89%. Regarding clinical trials of new treatments, two lines of evidence point to the low replication validity of initial studies. First, only 40% of phase II trials claiming a significant benefit of new drugs are confirmed when those drugs are subsequently tested in phase III trials [6, 7]. Second, most initial studies of a new treatment report a larger effect than subsequent studies or meta-analyses on the same topic [8–10, 11, 12]. Similarly, initial studies reporting a significant genetic association with various pathologies are often shown to report inflated effects when compared to the results of subsequent studies [13–15]. When the effect sizes associated with biomarkers reported in 35 highly cited studies were compared to those reported in the corresponding meta-analyses [16], 14 were initial studies reporting a significant association, of which one was disconfirmed by the corresponding meta-analysis and seven reported an effect size at least twice as large as that the one indicated in the corresponding meta-analysis.

The methods used in the four studies listed above were different. Two selected initial studies, and then searched for subsequent studies on the same topic [14, 16]. The other two exploited a database of 36 meta-analyses [13] that was extended to 55 meta-analyses [15]. Meta-analyses provide a unique way to compare initial findings with subsequent studies, and this approach has been used by Ioannidis and colleagues to estimate the replication validity of clinical trials [8–11]. Unfortunately, this approach cannot yet be widely used for preclinical studies because only a few meta-analyses exist in this area. In contrast, numerous recent meta-analyses have tested the association between markers or risk factors and various diseases. Because this type of biomedical research is crucial for improving the understanding and diagnosis of diseases, an estimate of the replication validity of these studies is timely. Since the four previously published estimates [13–16] mainly focused on genetic associations and were based on a relatively small number of meta-analyses, we decided to investigate this question on a larger scale.

According to Button and colleagues [17], “the average statistical power of neuroscience studies is very low. The consequences of this include overestimates of effect size and low reproducibility of the results.” Their analysis used a database of 49 meta-analyses published in 2011 in the field of neuroscience, most pertaining to psychiatric disorders or neurological diseases. This raises two questions. First, is the replication rate of initial studies in psychiatry similar to neurology? Second, how do these compare to a sample of somatic diseases? The present study was designed to address both questions. To assess the credibility of studies investigating the association of markers or risk factors with diseases, we collected all meta-analyses published

during a five year period related to four psychiatric disorders (attention deficit hyperactivity disorder (ADHD), autism, unipolar depression, schizophrenia), four neurological diseases (Alzheimer and Parkinson diseases, epilepsy and multiple sclerosis) and four somatic diseases (breast cancer, glaucoma, psoriasis and rheumatoid arthritis). We aimed to test whether initial studies were in agreement with the corresponding meta-analyses and to calculate their effect size inflation. We then explored several factors that might influence discrepancies between initial studies and meta-analyses: the level of statistical significance of the corresponding meta-analyses and of the initial studies, the biomedical domain, the study type (e.g., genetic), the sample size, the journal Impact Factor and the publication year.

Methods

Selection of meta-analyses for Inclusion

We searched PubMed for relevant articles, with the search limited to articles referenced as "meta-analysis" and published in English between January 1st 2008 and December 31st 2012. Studies published online in 2012 but printed in 2013 or 2014 were also considered. We used the following key words for each disease: "attention deficit hyperactivity disorder", "autism", "major depressive disorder", "epilepsy", "Alzheimer disease", "Parkinson disease", "multiple sclerosis", "breast cancer", "glaucoma", "psoriasis" and "rheumatoid arthritis".

Meta-analyses captured by this search strategy were screened by two authors (EDM and FG) for eligibility. First, the title and the abstract were considered. Second, full texts were obtained for the remaining studies and screened by two authors (EDM or FG). Articles were excluded if: 1) they were related to the treatment, screening or diagnosis of the disease, 2) if they reported on voxel-based brain imaging studies, 3) the disease itself was the risk factor for another outcome, 4) no meta-analysis was conducted, or 5) the article was not related or relevant to the disease. Meta-analyses were included if: 1) they compiled results from at least 7 distinct datasets reported in at least 4 independent publications, 2) the effect size was expressed as mean difference, standardized mean difference (Cohen's *d*), odds ratio (OR) or risk ratio (RR), and 3) parameters were measured and/or validated by professionals (i.e., psychological traits that were only self-assessed by subjects were not considered).

A number of articles reported several meta-analyses. We included all of them providing that they were dealing with distinct parameters or risk factors. If an article reported several meta-analyses on the same parameter and the same outcome, we selected the most comprehensive one (i.e., the one containing the most datasets) and the one reporting the fixed effects rather than the random effects model. When two or more articles dealt with the same parameter and the same outcome, the most recent one was used. For genetic association studies, we used the meta-analysis results for the allelic model when the data were available. However, when the corresponding initial study put forward another model we selected this one.

Selection of initial studies and largest studies

From each meta-analysis we identified the initial study and we collected its full text article. We checked them to assert that no study on the same topic was previously published. We thus identified 110 initial studies that were not listed in the corresponding meta-analyses. These omissions were mainly due to two reasons. First, a meta-analysis article of 36 associations between brain imaging observations and schizophrenia explicitly mentioned that data published before 1998 were not taken into account. Therefore, we looked for the "true" initial studies. Second, initial studies were often replicated by the same group using an extended population and the second publication often included the initial data. Therefore, the corresponding meta-analyses did not include the initial study. Nevertheless, we used here the data of the

“true” initial studies. From each meta-analysis we also identified the largest study as the one that included the largest number of patients.

Data extraction

Data were extracted independently by two authors (EDM and FG), and discrepancies resolved by mutual consent. A third author (KB) conducted a further 10% data check, which resulted in a small number of errors being detected and corrected. The following data were extracted from initial, largest and meta-analysis articles: year of publication, first author, 2012 Impact Factor of the journal that published it, nominal statistical significance (based on $p < 0.05$), effect size, its p-value for datasets reporting a significant effect or its confidence interval for non-significant ones, number of patients and of healthy controls. Moreover, regarding meta-analyses we also extracted the number of included datasets.

If the effect size was reported as a mean difference, it was converted to Cohen's d using the software OpenMetaAnalyst [18]. Meta-analyses and their corresponding initial and largest studies were excluded if data were incomplete or if the reporting was unclear. In order to compare the effect sizes across all meta-analyses and diseases, effect sizes were natural logarithm transformed. The effect sizes expressed as Cohen's d were converted to $\ln(\text{OR})$ using the formula $\ln(\text{OR}) = d \times \pi / \sqrt{3}$. For each disease, we calculated the average effect size of the initial and largest studies as well as of the meta-analyses and the median of the number of patients. These raw data are given in Supporting Information (S1 File).

Sensitivity and specificity analysis

In order to describe the nominal agreement between each initial study and its corresponding meta-analysis we used, as previous authors have done [15], two concepts borrowed from biomedical diagnostic testing: sensitivity and specificity. In the context of our study, sensitivity expresses the conditional probability that an initial study predicts a significant effect when a significant effect is actually confirmed by the corresponding meta-analysis. Specificity expresses the probability that an initial study reporting a non-significant effect corresponds to a meta-analysis supporting the same view.

This agreement was coded according to four possibilities: 1) initial study and meta-analysis both report a significant effect at $p < 0.05$, 2) initial study and meta-analysis both report a non-significant effect, 3) the initial study reports a significant effect whereas its associated meta-analysis does not (false positive), and 4) the initial study reports a non-significant effect whereas the associated meta-analysis reports a significant one (false negative). In addition we observed that six initial studies reporting a significant effect corresponded to meta-analyses reporting a significant effect in the opposite direction. These six initial studies were coded as case 3 (false positive). The sensitivity and specificity were calculated using the Clinical Calculator 1 (vassarstats.net). Their confidence intervals at 95% were given by this online calculator according to Newcombe [19].

Defining a small effect

We used our dataset of meta-analyses reporting a significant association to define a small effect in each of our four association types: cognitive/behavioral, brain imaging, genetic and other. For each category we sorted in quintiles the significant effect sizes. The largest effect size of the lowest quintile was defined as the maximal value of a small effect. All effects sizes of the 65 significant cognitive/behavioral studies were expressed as standardized mean difference and a small effect was estimated to be $d \leq 0.51$. All but two effect sizes of the 69 significant brain imaging studies were expressed as standardized mean difference. The two others were expressed as odds ratio and were converted to standardized mean difference. This gave a small

effect as $d \leq 0.29$. All 136 significant genetic associations were expressed as odds ratio and a small effect was defined as $OR \leq 1.11$. Finally, the 146 significant “other” studies were first divided into two categories: 60 were expressed as standardized mean difference and 96 as odds ratios. Regarding the former, a small effect was $d \leq 0.30$ and the latter was $OR \leq 1.20$.

Statistical analyses

Effect sizes of initial or largest studies were compared to those of the corresponding meta-analyses using Wilcoxon’s test for paired observations, as others have done with similar data [20, 21]. Other comparisons were tested using the χ^2 test. The relationship between effect size inflation and Impact factor was analyzed using linear regression and also Spearman’s rank correlation. These statistical analyses were done using SigmaPlot software.

Results

Characteristics of included studies

We included 663 meta-analyses reporting an association between a marker or a risk factor and a disease outcome in one of the three domains of interest (psychiatry, neurology and somatic diseases). Among these, 66 meta-analyses were related to cognitive test or behavioral observation and 97 to brain imaging studies. All but two meta-analyses were in the domain of psychiatry (see Table 1). Genetic studies were widely represented in all three domains (Table 1). Finally, associations classified as “other” corresponded to a wide variety of studies (e.g., epidemiology, blood tests, physical characteristics). Table 1 summarizes the number of articles selected at each step of the process, and the number of meta-analyses included for each pathology. The selection of studies is also shown in flow diagrams (see supporting information S1 Fig).

Do initial studies agree with subsequent meta-analyses?

As previous authors have done [15], we calculated the sensitivity and specificity of initial studies to predict the presence or absence of an association indicated by the corresponding meta-analysis (defined as statistically significant at $p < 0.05$). The sensitivity of initial studies ranged from 0.58 to 0.64 and was very similar across the three biomedical domains (Table 2). In contrast, the specificity ranged from 0.47 to 0.70 (Table 2) and there was some evidence that this was lower for initial studies in psychiatry when compared to neurology (χ^2 test: $p = 0.0055$) and somatic diseases ($p = 0.059$). We used a ROC diagram to plot the sensitivity as a function of (1-specificity). This indicated that the agreement between initial studies and corresponding meta-analyses may not be better than chance in psychiatry whereas it was somewhat better in neurology and somatic diseases (Fig 1).

Previous studies have shown that meta-analyses reporting a significant effect at $p < 0.05$ are sometimes disconfirmed by subsequent meta-analyses whereas those reporting stronger statistical evidence ($p < 0.005$) were rarely disconfirmed [20, 22]. This higher level of statistical stringency may increase the credibility of scientific research [23]. Furthermore, a meta-analysis reporting a non-significant effect only means that its confidence interval contains the null value, but does not prove the absence of an association. However, the confidence interval can be used to reject an effect larger than a predefined threshold [24]. If the 95% confidence interval excludes anything other than a small effect, the study under consideration is able to reject a larger effect. However, there is no general rule that defines what is a small effect—it depends on the type of association [22]. Here we used our database of 425 significant associations reported by meta-analyses to define a small effect in each of our four association types: cognitive/behavioral, brain imaging, genetic and other (see Methods).

Table 1. Number of studies identified in three research domains.

	Psychiatric disorders				Neurological diseases				Somatic diseases			
	ADHD	ASD	MDD	SCZ	AD	Epi	MS	PD	BC	Glau	Pso	RA
PubMed search	118	71	553	454	197	147	137	139	811	75	81	285
Examined full texts	43	36	114	198	117	29	44	74	345	19	22	82
Included articles	20	13	28	59	35	12	18	20	89	9	8	22
Included meta-analyses	40	24	53	203	50	15	37	57	114	21	15	34
<i>Cognitive/behavioral</i>	14	0	9	43	1	0	0	1	0	0	0	0
<i>Brain imaging</i>	2	10	19	64	1	1	0	0	0	0	0	0
<i>Genetic studies</i>	12	4	8	56	33	3	17	30	61	17	12	31
<i>Other studies</i>	12	10	17	40	15	11	20	26	53	4	3	3

ADHD: attention deficit hyperactivity disorder. ASD: autism spectrum disorder. MDD: major depressive disorder. SCZ: schizophrenia. AD: Alzheimer disease. Epi: epilepsy. MS: multiple sclerosis. PD: Parkinson disease. BC: breast cancer. Glau: glaucoma. Pso: psoriasis. RA: rheumatoid arthritis.

doi:10.1371/journal.pone.0158064.t001

We therefore conducted a sensitivity analysis where we restricted our analysis to associations investigated by stringent meta-analyses. We defined a meta-analysis as stringent either if it reported an effect as significant at $p < 0.005$ or, when non-significant, if its confidence interval excluded a threshold effect, as determined relatively to typical effect sizes in that field. Using these criteria we categorized 385 meta-analyses as stringent, of which 326 reported a significant effect (Table 2). When initial studies were compared to these corresponding stringent meta-analyses, sensitivity and specificity remained similar to those observed when all 663 associations were considered (Table 2). Again, the ROC diagram shows that the agreement between initial studies in psychiatry and their corresponding stringent meta-analyses might not be better than chance (Fig 1).

Do initial studies report inflated effect sizes?

Even if initial studies and corresponding meta-analyses both report a significant effect, initial studies might be misleading by reporting an inflated estimate of the effect size. Among the 256

Table 2. Diagnostic performance of initial studies against the statistical significance of the corresponding meta-analyses.

	PSY	NEURO	SOMA
All meta-analyses	320	159	184
significant meta-analyses	221 (69.1%)	100 (62.9%)	104 (56.5%)
sensitivity	0.612 (0.544 0.676)	0.636 (0.533 0.729)	0.584 (0.482 0.680)
specificity	0.475 (0.376 0.577)	0.700 (0.566 0.808)	0.614 (0.501 0.717)
Stringent meta-analyses	205	85	95
significant meta-analyses	178 (86.8%)	77 (90.6%)	71 (74.7%)
sensitivity	0.625 (0.549 0.696)	0.649 (0.531 0.752)	0.718 (0.597 0.816)
specificity	0.533 (0.346 0.712)	0.875 (0.477 0.993)	0.625 (0.408 0.804)

PSY: psychiatric disorders; NEURO: neurological diseases; SOMA: somatic diseases. Confidence intervals at 95% are given in parentheses.

doi:10.1371/journal.pone.0158064.t002

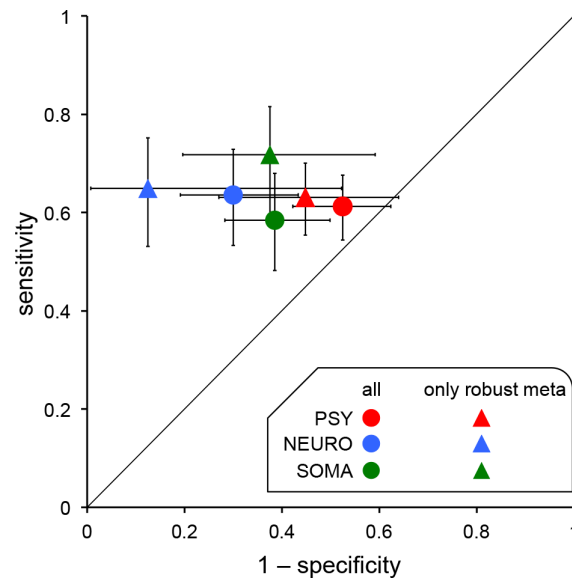


Fig 1. ROC diagram expressing the sensitivity and specificity of initial studies in agreeing with corresponding meta-analyses regarding the presence or absence of a nominally significant effect. For the three biomedical domains the diagram either considered all 663 pairs of initial studies and meta-analyses (circles) or only the 385 initial studies paired with stringent meta-analyses (triangles). Error bars represent the confidence intervals at 95%. The dashed diagonal line corresponds to no discrimination above chance. The fact that errors bars related to psychiatry cross this line suggests that the agreement between initial studies and corresponding meta-analyses might not be better than chance. PSY: psychiatry; NEURO: neurology; SOMA: somatic diseases.

doi:10.1371/journal.pone.0158064.g001

initial studies reporting a significant effect and paired with a meta-analysis reporting a significant effect in the same direction, 150 initial studies (59%) reported an effect size inflated by more than 50%, of which 97 were inflated by more than 100%. When we analyzed the 12 pathologies under consideration separately, the average effect size reported by initial studies was always larger than that reported by subsequent meta-analyses (Fig 2A). In contrast the effect size reported by the corresponding largest studies were much more consistent with the meta-analyses (Fig 2B and 2D). Indeed, for all pathologies, the effect size distribution differed between those reported by initial studies and those reported by meta-analyses (Wilcoxon’s test for paired observations). There was clear statistical evidence for most of these differences ($p < 0.002$) except for autism ($p = 0.025$), epilepsy ($p = 0.02$) and psoriasis ($p = 0.009$). A similar pattern was observed when considering only initial studies reporting a significant effect (Fig 2C).

Are initial studies consistent with subsequent meta-analyses?

To summarize the consistency between initial studies and corresponding meta-analyses we classified as replicated: *i*) initial studies reporting a non-significant effect and corresponding to non-significant meta-analyses, and *ii*) initial studies reporting a significant effect and corresponding to significant meta-analyses providing that the inflation of the effect size did not exceed a certain percentage threshold. We tested two inflation thresholds: 100% and 50% (Table 3). Regardless of the threshold, initial studies dealing with psychiatric disorders were less often in agreement with the corresponding meta-analyses than those dealing with neurological diseases (Table 3). However, there was no evidence that replication rates differed between psychiatric disorders and somatic diseases (Table 3). We also analyzed initial studies reporting a significant effect only and we did not observe any evidence of a difference between the three biomedical domains (Table 3).

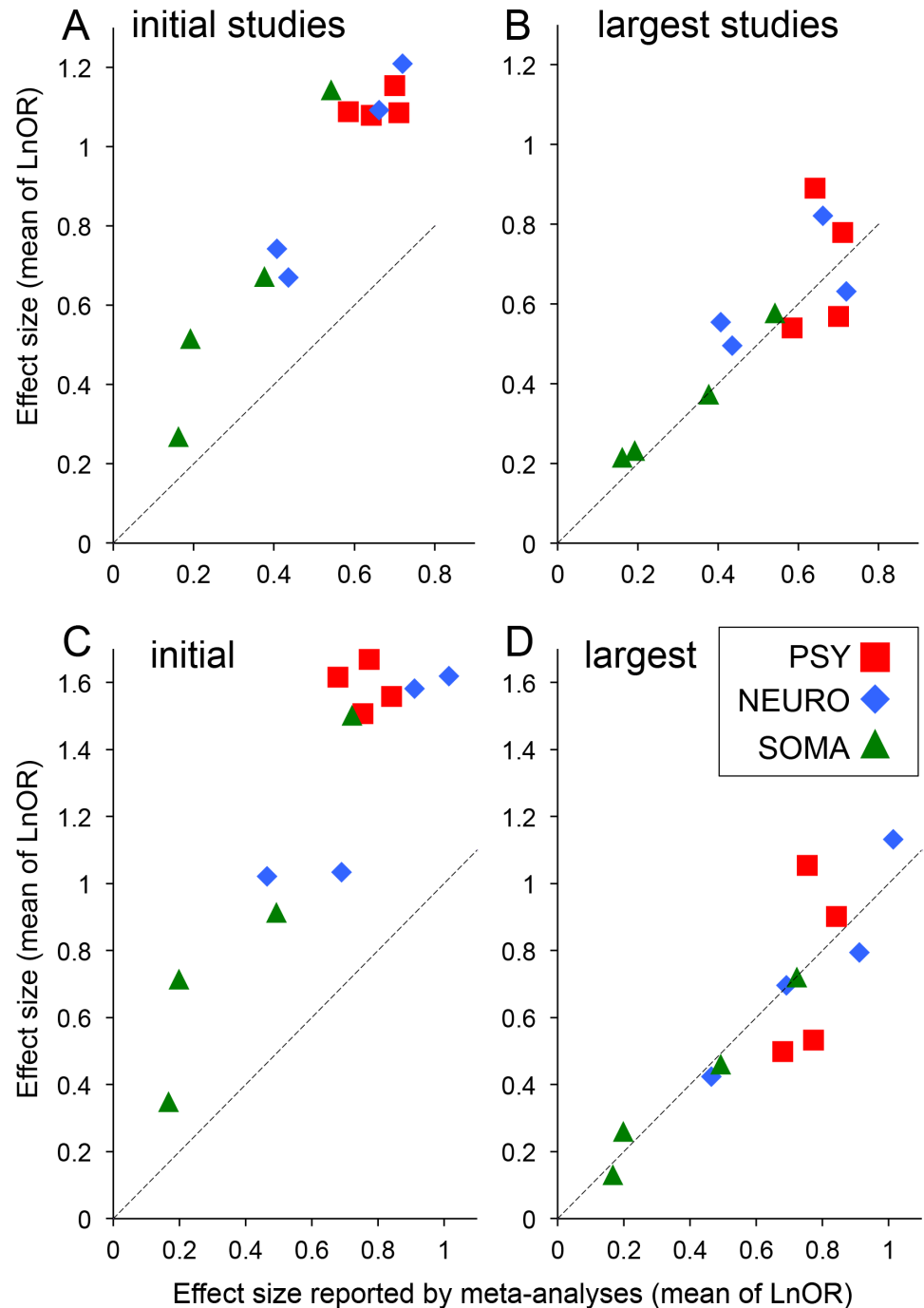


Fig 2. Comparison between effect sizes reported by initial or largest studies and their corresponding meta-analyses. All effect sizes were expressed as LnOR (standardized mean differences were converted to LnOR as described in the Methods). (A) and (C) Mean of the effect sizes reported by initial studies for each pathology. (B) and (D) Mean of the effect sizes reported by largest studies for each pathology. (A) and (B) All 663 trios were considered. (C) and (D) Only the 359 initial studies reporting a significant effect and their corresponding largest studies. The dashed lines correspond to equality between the effect sizes of the initial studies or of the largest studies and the effect sizes reported by the meta-analyses. PSY: psychiatry; NEURO: neurology; SOMA: somatic diseases.

doi:10.1371/journal.pone.0158064.g002

Table 3. Number of initial studies consistent with corresponding meta-analyses.

	number of studies						comparisons		
	PSY		NEURO		SOMA		PSY /	PSY /	NEURO /
	#	%	#	%	#	%	NEURO	SOMA	SOMA
All initial studies	320		159		184				
initial consistent (inflation ≤100%)	131	40.9	85	53.5	84	45.7	p = 0.0094	p = 0.30	p = 0.15
initial consistent (inflation ≤50%)	101	31.6	71	44.7	75	40.8	p = 0.0049	p = 0.037	p = 0.47
<i>Only initial reported as significant</i>									
number of studies	187		81		91				
initial consistent (inflation ≤100%)	83	44.4	43	53.1	33	36.3	p = 0.19	p = 0.20	p = 0.27
Only highly significant initial									
number of studies with p < 0.005	94		44		48				
initial consistent (inflation ≤100%)	42	44.7	23	52.3	23	47.9	p = 0.41	p = 0.71	p = 0.68
Without cognitive/behavioral	254		157		184				
initial consistent (inflation ≤100%)	90	35.4	84	53.5	84	45.7	p = 0.0003	p = 0.031	p = 0.15
initial consistent (inflation ≤50%)	70	27.6	71	45.2	75	40.8	p = 0.0003	p = 0.0039	p = 0.41
<i>Only initial reported as significant</i>									
number of studies	142		79		91				
initial consistent (inflation ≤100%)	45	31.7	42	53.2	33	36.3	p = 0.0019	p = 0.47	p = 0.028

PSY: psychiatric disorders; NEURO: neurological diseases; SOMA: somatic diseases.

doi:10.1371/journal.pone.0158064.t003

Table 4. Number of initial studies consistent with corresponding stringent meta-analyses.

	number of studies						comparisons		
	PSY		NEURO		SOMA		PSY / NEURO	PSY / SOMA	NEURO / SOMA
	#	%	#	%	#	%			
All initial studies	205		85		95				
initial consistent (inflation $\leq 100\%$)	93	45.4	44	51.8	47	49.5	p = 0.32	p = 0.51	p = 0.76
<i>Only initial reported as significant</i>									
number of studies	124		51		60				
initial consistent (inflation $\leq 100\%$)	77	62.1	37	72.5	32	53.3	p = 0.19	p = 0.26	p = 0.037

PSY: psychiatric disorders; NEURO: neurological diseases; SOMA: somatic diseases.

doi:10.1371/journal.pone.0158064.t004

When considering only initial studies paired with 385 stringent meta-analyses, these results did not change substantially (Table 4). However, when considering only initial studies reporting a significant effect, the percentage of initial studies in agreement with corresponding stringent meta-analyses was larger than those observed when considering all 663 pairs (see Table 4 vs Table 3). This is due to the fact that the percentage of meta-analyses reporting a non-significant effect was much smaller when we only considered stringent meta-analyses (Table 2). Therefore, the number of significant initial studies associated with non-significant stringent meta-analyses was reduced.

Are highly significant initial studies more reliable?

We applied our criteria of a stringent study to initial studies by removing those reporting a significant effect with a p-value >0.005 and those that did not exclude a small effect on the basis of their confidence interval. On this basis only five of the 304 initial studies reporting a non-significant effect were able to exclude a small effect and only one was consistent with the corresponding meta-analysis. Of the 359 initial studies reporting a significant effect, only 186 reported a p-value <0.005 . We compared their replication rate to that of all initial studies reporting a significant effect (Table 3). We observed no difference. Therefore, initial studies reporting a highly significant effect ($p < 0.005$) are no more reliable than those reporting a p-value between 0.05 and 0.005.

Does replication rate of initial studies differ by study type?

We tested whether the agreement of initial studies with their corresponding meta-analyses depends on the biomedical domain and on the type of study. In order to do so, we defined a replication rate as the percentage of initial studies that were in nominal agreement with meta-analyses and whose effect size inflation was $\leq 100\%$. In this subgroup analysis, we did not take into account four associations with neurological diseases (two brain imaging and two cognitive/behavioral studies) due to the small number of studies. We thus considered eight homogeneous subgroups. Associations with psychiatric disorders were sorted into four subgroups:

Table 5. Replication rate of initial studies in 8 subgroups.

	All meta-analyses			Stringent meta-analyses		
	# stud	# consist	%	# stud	# consist	%
Psychiatry						
Cognitive/behavioral	66	41	62.1	65	41	63.1
Brain imaging	95	36	37.9	55	23	41.8
Genetic	80	26	32.5	26	7	26.9
“Other” studies	79	28	35.4	59	22	37.3
Neurology						
Genetic	83	40	48.2	39	16	41.0
“Other” associations	72	42	58.3	43	26	60.5
Somatic diseases						
Genetic	121	60	49.6	57	32	56.1
“Other” associations	63	24	38.1	39	16	41.0

Initial studies are consistent if inflated by 100% or less.

doi:10.1371/journal.pone.0158064.t005

cognitive/behavioral, brain imaging, genetic and “other” studies. Associations with neurological and somatic diseases were both sorted into two subgroups: genetic and “other” studies. [Table 5](#) gives the number and percentage of initial studies that were consistent with their corresponding meta-analyses in each subgroup, whether considering all 659 meta-analyses or only the 382 defined as stringent. Data restricted to initial studies reporting a significant effect are given in [Table 6](#).

Tables [5](#) and [6](#) show that the replication rate of initial studies differs considerably between subgroups. Within psychiatry, initial studies of cognitive/behavioral associations exhibited the highest replication rate, whereas those regarding genetic associations were the least replicated. As all but two cognitive/behavioral meta-analyses were related to psychiatric disorders, we also

Table 6. Replication rate of significant initial studies in 8 subgroups.

	All meta-analyses			Stringent meta-analyses		
	# stud	# consist	%	# stud	# consist	%
Psychiatry						
Cognitive/behavioral	45	38	84.4	44	38	86.4
Brain imaging	50	21	42.0	28	18	64.3
Genetic	46	3	6.5	16	1	6.3
“Other” studies	46	21	45.7	36	20	55.6
Neurology						
Genetic	37	14	37.8	19	11	57.9
“Other” associations	41	27	65.9	29	24	82.8
Somatic diseases						
Genetic	61	24	39.3	40	24	60.0
“Other” associations	30	9	30.0	20	8	40.0

Initial studies are consistent if inflated by 100% or less.

doi:10.1371/journal.pone.0158064.t006

compared the replication rate of initial studies in the three domains when restricting the analysis to biomarkers and biological risk factors (Table 3). This analysis shows that initial studies in biological psychiatry were less reliable than those of both other biomedical domains.

Does replication rate depend on sample size and “true” effect size?

In order to elucidate the factors contributing to the replication rate of initial studies, we built three-dimensional graphs expressing for each subgroup how the replication rate was related to the “true” effect size (i.e., the effect size reported by stringent meta-analyses) and to sample size (i.e., the number of patients in initial studies) (Fig 3). Whether considering all initial studies

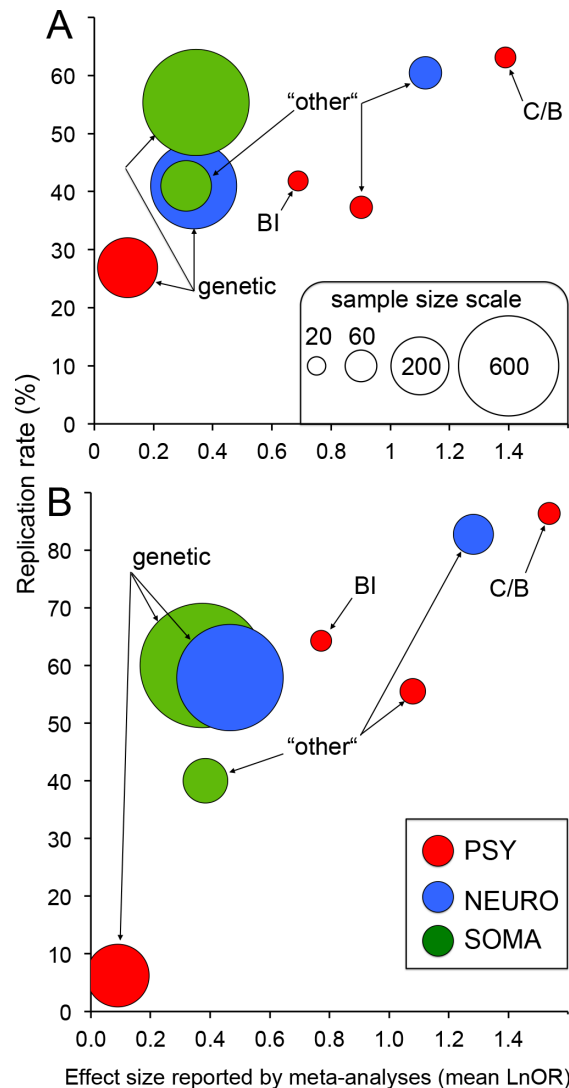


Fig 3. Replication rate of initial studies depending on the “true” effect and on the sample size. The replication rate was calculated for the eight subgroups as the percentage of initial studies that were consistent with meta-analyses and whose effect size inflation was $\leq 100\%$ (raw data are given in Tables 5 and 6). The average of the “true” effect size was the mean of the summary effect size reported by each stringent meta-analysis for each subgroup. The sample size corresponded to the median of the number of patients of all initial studies in each subgroup and was represented as the area of the circles (same scale for (A) and (B)). (A) All initial studies paired with stringent meta-analyses ($n = 382$). (B) Only initial studies reporting a significant effect and paired with stringent meta-analyses ($n = 232$). BI: brain imaging studies; C/B: cognitive/behavioral studies; PSY: psychiatric disorders; NEURO: neurological diseases; SOMA: somatic diseases.

doi:10.1371/journal.pone.0158064.g003

(Fig 3A) or only those reporting a significant effect (Fig 3B), subgroups exhibiting the lowest replication rates were associated with either a smaller “true” effect size or a lower sample size or both.

Are largest studies more reliable?

We used the same procedures to evaluate the performance of largest studies against their corresponding meta-analyses. With respect to the nominal agreement with corresponding meta-analyses, the sensitivity of largest studies (Table 7) was similar to that of initial studies (Table 2), but their specificity was better. This was true either when all 663 associations were considered or when the analysis was restricted to stringent meta-analyses (Table 7). Moreover, there was no evidence that the performance of the largest studies differed between the three domains (Table 7).

The replication rate of largest studies was much better than that of initial studies. It ranged from 62.5% to 67.9% when all largest studies were considered, and from 70.6% to 80.4% when only significant largest studies were taken into account. Again, we observed no significant difference between the three domains.

Are studies published in high Impact Factor journals more reliable?

In our sample of 663 associations, initial studies dealing with psychiatric disorders were published in scientific journals exhibiting a higher Impact Factor than those dealing with neurological diseases or with our set of somatic diseases (Impact Factor median: 7.7 versus 4.8 versus 5.6, respectively). However, there was no statistical evidence that the Impact Factor distributions differed between the three domains.

To test the association of Impact Factor and the replication rate of initial studies, we calculated an inflation index for every initial study defined as the ratio of its effect size divided by the effect size reported by the corresponding meta-analysis [25]. We plotted this ratio against the Impact Factor of the journal that published the initial study. Data were analyzed using linear regression. This analysis was performed separately for each three domains and for the eight subgroups. No negative relationships were observed. We found no evidence of a positive relationship between inflation score and Impact Factor except in three subgroups: genetic studies related to psychiatry ($r^2 = 0.06$, $p = 0.026$, $n = 80$) and to neurology ($r^2 = 0.06$, $p = 0.025$, $n = 83$) as well as “other” studies related to somatic diseases ($r^2 = 0.11$, $p = 0.008$, $n = 63$). Nevertheless, the strength of association in each case was modest. We also tested whether the inflation index was correlated with the Impact factor using Spearman’s rank correlation. None of the eleven tested associations reached statistical significance at $p < 0.05$.

Table 7. Diagnostic performance of largest studies against the statistical significance of the corresponding meta-analyses.

	PSY	NEURO	SOMA
All meta-analyses	320	159	184
sensitivity	0.586 (0.518 0.652)	0.646 (0.543 0.738)	0.538 (0.438 0.636)
specificity	0.82 (0.728 0.887)	0.817 (0.691 0.901)	0.85 (0.749 0.917)
Stringent meta-analyses	205	85	95
sensitivity	0.669 (0.594 0.736)	0.795 (0.686 0.875)	0.643 (0.519 0.751)
specificity	0.926 (0.742 0.987)	1 (0.598 1)	0.875 (0.665 0.967)

PSY: psychiatric disorders; NEURO: neurological diseases; SOMA: somatic diseases. Confidence intervals at 95% are given in parentheses.

doi:10.1371/journal.pone.0158064.t007

Does publication year influence the reliability?

In our sample of 663 initial studies only 40 were published before 1980 and none in 2010 or later. Therefore, in order to test a possible influence of the publication year we sorted our sample into four groups, those published before 1980 and those published during the last three decades (1980s, 1990s and 2000s). There was no evidence that publication year was associated with the replication rate of initial studies whether analyzing all associations together or considering each three domains separately.

However, among genetic studies, the publication year of initial studies reporting a significant effect deserves to be examined in detail. Indeed, in our sample of initial studies, only 2/46 psychiatric studies were published between 2005 and 2009 whereas for neurological and somatic diseases 17/37 and 22/61 studies were published during this five-year period, respectively. Because for neurological and somatic subgroups the sample size increased more than six-fold during this period, due to the growth in large collaborative studies, the replication rate of their initial studies was three times larger than for those published before 2005. Thus, the fact that our sample of initial genetic studies related to psychiatry included only two studies published after 2004 partly contributes to the overall impression that they were less often replicated (Table 6).

Are initial studies representative of later studies?

Fig 4 shows the relationship between the number of patients in initial studies and the average number of patients in subsequent ones. This average number of patients was calculated by subtracting the number of patients in initial studies from the total number of patients included in each meta-analysis and by dividing it by the number of datasets minus one. For the eight subgroups, the sample size of initial studies was always smaller than for the averaged subsequent

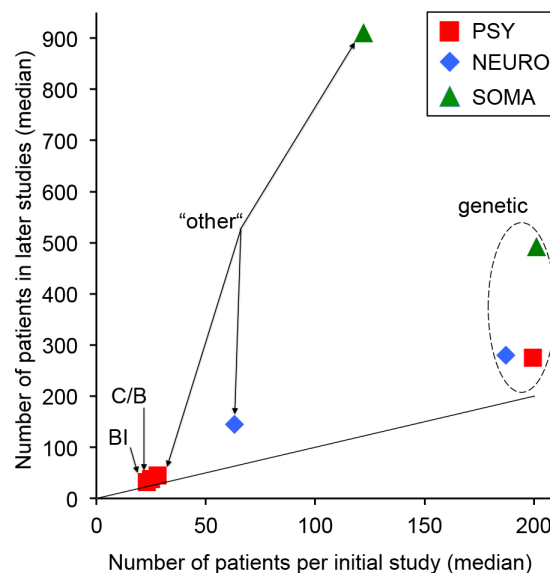


Fig 4. Relationship between the number of patients in initial studies and the average number of patients in subsequent studies. The average number of patients in subsequent studies was calculated by subtracting the number of patients in initial studies from the total number of patients included in each meta-analysis and by dividing it by the number of datasets minus one. The 659 averaged sample sizes of subsequent studies were sorted in eight subgroups and we calculated the median for each subgroup. These eight medians were plotted as a function of the medians of the number of patients in initial studies for each subgroup. The dashed line corresponds to equality between both median types. BI: brain imaging studies; C/B: cognitive/behavioral studies.

doi:10.1371/journal.pone.0158064.g004

studies, but with noticeable differences (Fig 4). The average sample size of subsequent studies was at least twice as large as for initial studies regarding three subgroups: somatic studies in genetics as well as neurological and somatic studies of the “other” type (ratio: 2.4, 2.3 and 7.5, respectively). Because in these three subgroups the subsequent studies included, on average, many more patients, many are expected to be more reliable than initial studies. In contrast, concerning the five other subgroups (i.e., the four subgroups related to psychiatry and the genetic neurological studies), the number of patients in subsequent studies was, on average, only 1.4 to 1.6 times larger than that of the corresponding initial studies. Therefore, the replication rate of these subsequent studies would be expected to be similar to that of the corresponding initial studies. However, this warrants further research.

Comments

Our observations confirm that the replication validity of association studies is low. The agreement between initial studies and their corresponding meta-analyses regarding the presence or absence of a significant effect is not better or only poorly better than chance depending on the biomedical domain. Moreover, 43% of initial significant studies paired with a significant meta-analysis reported an effect size inflated by more than 100%. Our major finding is that lack of replication validity and effect size inflation differ across biomedical domains and study types. Among initial studies reporting a significant effect, the replication rate ranged from 6.3% for genetic studies of psychiatric disorders to 86.4% for cognitive/behavioral studies of psychiatric disorders. When comparing the three biomedical domains we observed no major differences, but when this comparison was restricted to biomarkers and biological risk factors, initial studies related to biological psychiatry appeared less reliable than those of both other domains.

Button and colleagues suggested that small sample size undermines the reliability of many neuroscience studies [17]. The present study reaches the same general conclusion and extends it. First, in our samples, initial neurological studies are at least as reliable as those related to a set of four somatic diseases. Second, cognitive/behavioral initial studies related to psychiatric disorders appeared much more reliable than other psychiatric studies. Third, initial studies related to biological psychiatry are less reliable than those related to neurological diseases.

Theoretical simulations show that small sample size undermines reproducibility and increases effect size inflation [1, 17, 26]. Previous observational studies [9, 11, 27] as well as the present study support this view. Moreover, the comparisons between eight subgroups illustrate that the relationship between sample size and effect size inflation is relative. When the “true” effect size is large, as in our sample of cognitive/behavioral studies, even initial studies with a small sample size are replicated. In contrast, initial genetic studies related to psychiatry were less often replicated than in neurology not because their sample size was smaller but because it was too small to accurately deal with a much smaller “true” effect size. The fact that the replication rate of largest studies is much better than that of initial studies further supports the view that inadequate sample size is a major cause of low replication validity.

Three previous studies have already highlighted that many brain imaging studies report inflated effect sizes [28–30]. In particular, Ioannidis evaluated the 461 primary datasets included in 41 meta-analyses on the associations between psychiatric disorders and brain volume abnormalities [28]. This demonstrated that the number of primary studies reporting a significant effect was about twice as large as what could be expected from the “true” effects reported by the meta-analyses. Our observations regarding the 95 brain imaging studies related to psychiatric disorders are consistent with this previous study.

The effect size inflation and the poor replication validity of candidate gene studies of complex diseases have been already described [13, 15, 31]. In particular Trikalinos and colleagues

analyzed 55 meta-analyses and showed that initial studies poorly predicted the presence or absence of a nominally significant effect reported by the corresponding meta-analysis [15]. More surprisingly, but in agreement with our own observations, they showed that this predictive value was not improved when considering only initial studies reporting a highly significant effect ($p < 0.005$).

In our sample, 38 out of 45 initial studies reporting a significant association between cognitive/behavioral observations and psychiatric disorders were in agreement with the corresponding meta-analysis. This represents a much better performance than expected. Indeed, collaborative attempts to replicate observations in cognitive psychology showed that only 21 out of 42 initial studies reporting a significant effect were replicated [21]. Moreover, psychological science suffers from the same reporting biases as biomedical science [32].

We observed weak positive relationships between the inflation score of initial studies and the Impact Factor of the journal that published them for three subgroups including genetic studies in psychiatry and neurology. No negative relationships were observed. Our observations are in agreement with the positive relationship previously reported regarding 81 primary genetic studies related to psychiatry, although our relationships are clearly weaker [25]. Nevertheless, both studies at least suggest that the reliability of initial studies does not increase with the journal Impact Factor and this conclusion has also been drawn about preclinical studies [33].

Limitations

Our study is based on the view that the summary effect reported by a meta-analysis is a reliable estimate of the “true” effect. For this reason we only considered meta-analyses dealing with at least seven independent datasets published in at least 4 independent publications. Indeed, a comparison between pairs of meta-analyses published at a five-years interval about the same association showed that the change in summary effect size was inversely related to the number of datasets [20]. Despite this selection criterion, it is likely that some meta-analyses of our sample still reported an inflated effect or even a false positive effect. Indeed, publication bias (i.e., the preferential publication of positive results) and the inclusion of primary studies with low methodological quality are likely to inflate the effect size estimated by meta-analyses [34, 35]. However, without study pre-registration, these biases are difficult to detect. Several tests have been proposed to identify publication bias in meta-analyses and to correct for it [34, 35, 36, 37], but these methods suffer from limitations and often require meticulous explicit modeling of each potential source of bias [37, 38]. Ioannidis and colleagues propose two methods to probe the robustness of meta-analyses that report a statistically significant effect at $p < 0.05$ [20, 22, 39]. Both methods require no assumption about the sources of potential biases, but are not easy to apply in the context of our large database. However, these studies inspired us to adopt a simpler approach to identify, among nominally significant meta-analyses, those likely to retain their statistical significance with the inclusion of further studies.

We hypothesized that meta-analyses reporting highly significant effects ($p < 0.005$) are more reliable than those reporting a p value between 0.05 and 0.005. This hypothesis is supported by theoretical considerations and by observational studies. Indeed, this more stringent p -value roughly corresponds to a Bayes factor of 50, which means that no more than one in 50 meta-analyses would report a false positive effect [23]. Moreover, Pereira and colleagues (2011) reported on 80 nominally significant ($p < 0.05$) meta-analyses of clinical trials published in 2005 that were updated in 2010 with additional trials. Among them, four lost their statistical significance in 2010 and all four had a p value between 0.005 and 0.05 in 2005. Likewise, using a Bayesian approach, Ioannidis (2008) tested the effect of formal statistical significance on the

credibility of 50 meta-analyses reporting a significant genetic association at $p < 0.05$ [22]. Among them, 27 had a p value below 0.005 of which 26 were analyzed as providing at least a strong support to the association. In contrast, none of the 23 genetic associations with a reported p value between 0.5 and 0.005 provided a strong support.

Ioannidis listed small sample size and small “true” effect size as two major causes of replication failure [1]. However, he also highlighted other causes related to the methodological quality of the studies and to the general characteristics of each research field (i.e., the number of associations tested, the influence of financial interests and the strength of the competition between researchers) [1]. However, because these other causes of replication failure are difficult to identify and quantify, especially when considering a large database, they were not investigated in the present study.

We collected and analyzed the data of 663 initial studies and their corresponding largest studies and meta-analyses. We did not individually consider each of the 10,154 primary datasets included in these 663 meta-analyses. Therefore, while we provide direct data about the replication rate of initial and largest studies, we have only indirect and limited evidence regarding that of other primary studies in each subgroup. Our emphasis on initial studies, and especially on those reporting a significant effect, is justified by their influence on subsequent research efforts [15], and by the fact that they are much more echoed by the media than subsequent studies [40].

Our selection of 12 pathologies is partly arbitrary. Regarding neurological diseases we excluded stroke because it is also a cardiovascular disease, traumatic brain injury because it is accidental and autism, which is rather a psychiatric disorder. Thus, our selection of neurological diseases includes all four severe diseases exhibiting the highest prevalence rate [41]. Our selection of four psychiatric disorders is less systematic: it included two severe disorders (autism and schizophrenia) and two less severe and more prevalent disorders (ADHD and unipolar depression). Finally, our selection of four somatic diseases is even more arbitrary. We selected them because their biological causes and risk factors are still poorly elucidated and because they cover a range of prevalence similar to that of both other domains.

Our sample of initial studies was extracted from meta-analyses built from at least seven primary datasets. We did not investigate initial studies not followed by meta-analyses or included in less extensive meta-analyses. Therefore, our sample represents a tiny fraction of all studies published during the last three decades about the association of a risk factor with one of our 12 pathologies. Indeed, among a random sample of 259 biomedical studies reporting observational data, only 6.2% were included in a systematic meta-analysis [42]. Whether our sample of initial studies was representative regarding replication validity remains to be elucidated.

Conclusion

The poor replication validity of initial biomedical studies has been put forward to recommend potential solutions that might improve it [4, 32, 43–45]. However, some of these recommendations might have unintended negative consequences [46]. Beside these recommendations, we would like to emphasize the opinion voiced by the Open Science Collaboration [21]: “If initial ideas were always correct, then there would hardly be a reason to conduct research in the first place. A healthy discipline will have many false starts as it confronts the limits of the present understanding”. Nevertheless, the differences between biological psychiatry, neurology and somatic diseases suggest that there is room for improvement at least in some subdomains.

An unrecognized consequence of the low reliability of initial studies is not related to the scientific process of uncertainty reduction *per se*, but to the media coverage of scientific findings. We showed with the case of ADHD that newspapers preferentially echo initial findings and

almost never inform the public when they are refuted or strongly attenuated although the majority of them are [40]. Therefore, the present estimates of the replication rates of initial studies in various domains and subdomains might be especially useful for journalists and for scientists when they interact with journalists. Initial findings should always be described as tentative, uncertain and requiring replication when reported in the scientific literature, the academic press releases and the media.

Supporting Information

S1 Fig. Twelve flow diagrams.

(DOCX)

S1 File. Raw data about initial studies, largest studies and meta-analyses for 12 pathologies.

(XLSX)

S1 Text. References of the 333 articles describing the 663 meta-analyses.

(DOCX)

S2 Text. References of the 485 articles describing the 663 initial observations.

(DOCX)

S3 Text. References of the 447 articles describing the 663 largest observations.

(DOCX)

Author Contributions

Conceived and designed the experiments: EDM KB TB MM FG. Performed the experiments: EDM KB FG. Analyzed the data: EDM TB FG. Contributed reagents/materials/analysis tools: TB MM FG. Wrote the paper: EDM KB TB MM FG.

References

1. Ioannidis JP. Why most published research findings are false. *PLoS Med.* 2005; 2(8):e124. PMID: [16060722](#)
2. Sarewitz D. Beware the creeping cracks of bias. *Nature.* 2012; 485(7397):149. doi: [10.1038/485149a](#) PMID: [22575922](#)
3. Munafo M, Noble S, Browne WJ, Brunner D, Button K, Ferreira J, et al. Scientific rigor and the art of motorcycle maintenance. *Nature biotechnology.* 2014; 32(9):871–3. doi: [10.1038/nbt.3004](#) PMID: [25203032](#)
4. Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. *Circulation research.* 2015; 116(1):116–26. doi: [10.1161/CIRCRESAHA.114.303819](#) PMID: [25552691](#)
5. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. *PLoS biology.* 2015; 13(6):e1002165. doi: [10.1371/journal.pbio.1002165](#) PMID: [26057340](#)
6. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. *Clinical pharmacology and therapeutics.* 2010; 87(3):272–7. doi: [10.1038/clpt.2009.295](#) PMID: [20130567](#)
7. Gan HK, You B, Pond GR, Chen EX. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. *Journal of the National Cancer Institute.* 2012; 104(8):590–8. doi: [10.1093/jnci/djs141](#) PMID: [22491345](#)
8. Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative meta-analyses. *Proceedings of the National Academy of Sciences of the United States of America.* 2001; 98(3):831–6. PMID: [11158556](#)
9. Trikalinos TA, Churchill R, Ferri M, Leucht S, Tuunainen A, Wahlbeck K, et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol.* 2004; 57(11):1124–30. PMID: [15612138](#)

10. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005; 294(2):218–28. PMID: [16014596](#)
11. Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. *JAMA*. 2012; 308(16):1676–84. doi: [10.1001/jama.2012.13444](#) PMID: [23093165](#)
12. Tajika A, Ogawa Y, Takeshima N, Hayasaka Y, Furukawa TA. Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up. *The British journal of psychiatry: the journal of mental science*. 2015; 207(4):357–62.
13. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet*. 2001; 29(3):306–9. PMID: [11600885](#)
14. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2002; 4(2):45–61.
15. Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG, Ioannidis JP. Establishment of genetic associations for complex diseases is independent of early study findings. *European journal of human genetics: EJHG*. 2004; 12(9):762–9. PMID: [15213707](#)
16. Ioannidis JP, Panagiotou OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA*. 2011; 305(21):2200–10. doi: [10.1001/jama.2011.713](#) PMID: [21632484](#)
17. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews Neuroscience*. 2013; 14(5):365–76. doi: [10.1038/nrn3475](#) PMID: [23571845](#)
18. Wallace BC, Schmid CH, Lau J, Trikalinos TA. Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. *BMC medical research methodology*. 2009; 9:80. doi: [10.1186/1471-2288-9-80](#) PMID: [19961608](#)
19. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine*. 1998; 17(8):857–72. PMID: [9595616](#)
20. Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *J Clin Epidemiol*. 2011; 64(10):1060–9. doi: [10.1016/j.jclinepi.2010.12.012](#) PMID: [21454050](#)
21. Open-Science-Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science (New York, NY)*. 2015; 349(6251):aac4716.
22. Ioannidis JP. Effect of formal statistical significance on the credibility of observational associations. *American journal of epidemiology*. 2008; 168(4):374–83; discussion 84–90. doi: [10.1093/aje/kwn156](#) PMID: [18611956](#)
23. Johnson VE. Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110(48):19313–7. doi: [10.1073/pnas.1313476110](#) PMID: [24218581](#)
24. Dienes Z. Using Bayes to get the most out of non-significant results. *Frontiers in psychology*. 2014; 5:781. doi: [10.3389/fpsyg.2014.00781](#) PMID: [25120503](#)
25. Munafo MR, Stothart G, Flint J. Bias in genetic association studies and impact factor. *Molecular psychiatry*. 2009; 14(2):119–20. doi: [10.1038/mp.2008.77](#) PMID: [19156153](#)
26. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology (Cambridge, Mass)*. 2008; 19(5):640–8.
27. Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. *Lancet (London, England)*. 2003; 361(9357):567–71.
28. Ioannidis JP. Excess significance bias in the literature on brain volume abnormalities. *Archives of general psychiatry*. 2011; 68(8):773–80. doi: [10.1001/archgenpsychiatry.2011.28](#) PMID: [21464342](#)
29. David SP, Ware JJ, Chu IM, Loftus PD, Fusar-Poli P, Radua J, et al. Potential reporting bias in fMRI studies of the brain. *PloS one*. 2013; 8(7):e70104. doi: [10.1371/journal.pone.0070104](#) PMID: [23936149](#)
30. Fusar-Poli P, Radua J, Frascarelli M, Mechelli A, Borgwardt S, Di Fabio F, et al. Evidence of reporting biases in voxel-based morphometry (VBM) studies of psychiatric and neurological disorders. *Human brain mapping*. 2014; 35(7):3052–65. doi: [10.1002/hbm.22384](#) PMID: [24123491](#)
31. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet*. 2003; 33(2):177–82. PMID: [12524541](#)
32. Ioannidis JP, Munafo MR, Fusar-Poli P, Nosek BA, David SP. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends in cognitive sciences*. 2014; 18(5):235–41. doi: [10.1016/j.tics.2014.02.010](#) PMID: [24656991](#)

33. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nature reviews Drug discovery*. 2011; 10(9):712. doi: [10.1038/nrd3439-c1](https://doi.org/10.1038/nrd3439-c1) PMID: [21892149](https://pubmed.ncbi.nlm.nih.gov/21892149/)
34. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical research ed)*. 1997; 315(7109):629–34.
35. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ (Clinical research ed)*. 2001; 323(7304):101–5.
36. Thompson S, Ekelund U, Jebb S, Lindroos AK, Mander A, Sharp S, et al. A proposed method of bias adjustment for meta-analyses of published observational studies. *International journal of epidemiology*. 2011; 40(3):765–77. doi: [10.1093/ije/dyq248](https://doi.org/10.1093/ije/dyq248) PMID: [21186183](https://pubmed.ncbi.nlm.nih.gov/21186183/)
37. Ioannidis JP. Commentary: Adjusting for bias: a user's guide to performing plastic surgery on meta-analyses of observational studies. *International journal of epidemiology*. 2011; 40(3):777–9. doi: [10.1093/ije/dyq265](https://doi.org/10.1093/ije/dyq265) PMID: [21233141](https://pubmed.ncbi.nlm.nih.gov/21233141/)
38. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Clinical research ed)*. 2011; 343:d4002.
39. Papatheodorou SI, Tsilidis KK, Evangelou E, Ioannidis JP. Application of credibility ceilings probes the robustness of meta-analyses of biomarkers and cancer risk. *J Clin Epidemiol*. 2015; 68(2):163–74. doi: [10.1016/j.jclinepi.2014.09.004](https://doi.org/10.1016/j.jclinepi.2014.09.004) PMID: [25433443](https://pubmed.ncbi.nlm.nih.gov/25433443/)
40. Gonon F, Kongsman JP, Cohen D, Boraud T. Why most biomedical findings echoed by newspapers turn out to be false: the case of Attention Deficit Hyperactivity Disorder. *PloS one*. 2012; 7(9):e44275. doi: [10.1371/journal.pone.0044275](https://doi.org/10.1371/journal.pone.0044275) PMID: [22984483](https://pubmed.ncbi.nlm.nih.gov/22984483/)
41. Hirtz D, Thurman DJ, Gwinn-Hardy K, Mohamed M, Chaudhuri AR, Zalutsky R. How common are the "common" neurologic disorders? *Neurology*. 2007; 68(5):326–37. PMID: [17261678](https://pubmed.ncbi.nlm.nih.gov/17261678/)
42. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS biology*. 2016; 14(1):e1002333. doi: [10.1371/journal.pbio.1002333](https://doi.org/10.1371/journal.pbio.1002333) PMID: [26726926](https://pubmed.ncbi.nlm.nih.gov/26726926/)
43. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet (London, England)*. 2014; 383(9913):267–76.
44. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. *Lancet (London, England)*. 2014; 383(9912):156–65.
45. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet (London, England)*. 2014; 383(9912):166–75.
46. Lewandowsky S, Bishop D. Research integrity: Don't let transparency damage science. *Nature*. 2016; 529(7587):459–61. doi: [10.1038/529459a](https://doi.org/10.1038/529459a) PMID: [26819029](https://pubmed.ncbi.nlm.nih.gov/26819029/)

Fig S1. Flow diagram ADHD

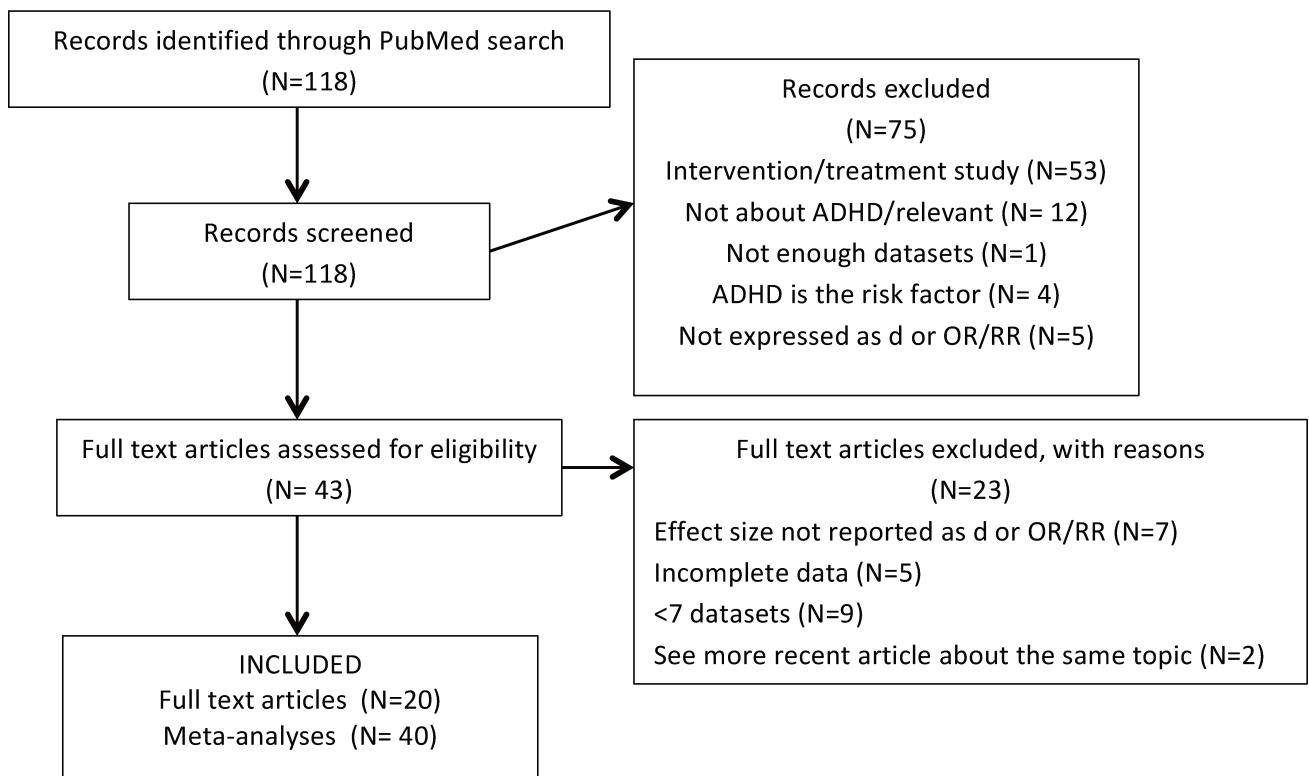


Fig S2. Flow diagram Autism (ASD)

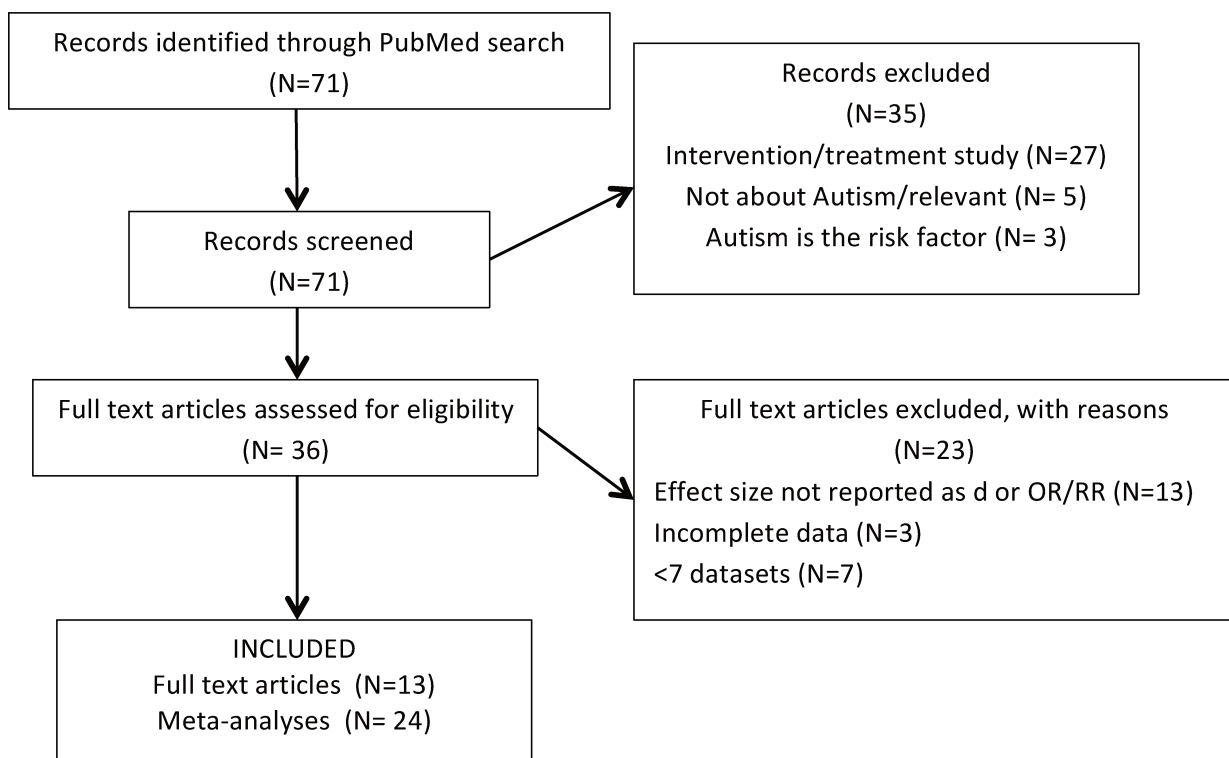


Fig S3. Flow diagram Major Depressive Disorder (MDD)

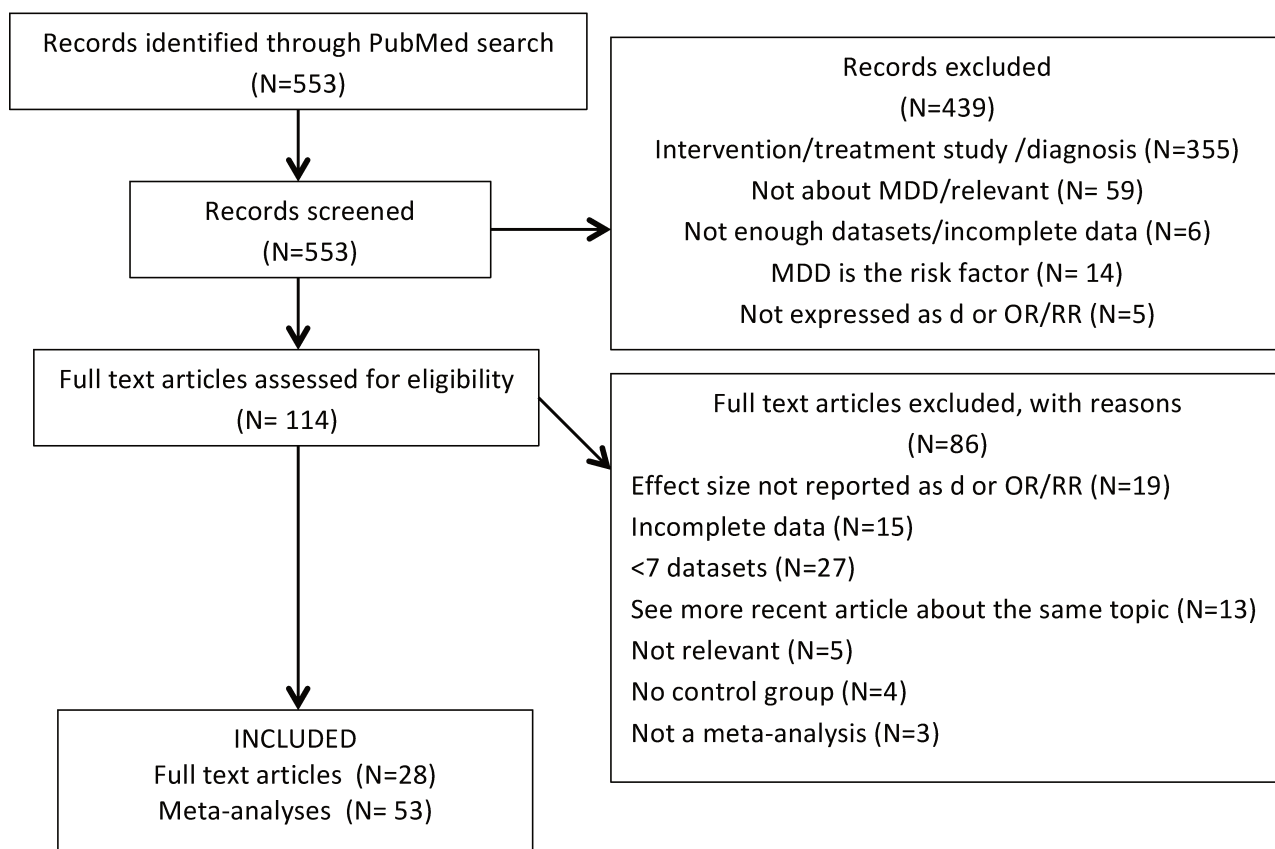


Fig S4. Flow diagram Schizophrenia (SCZ)

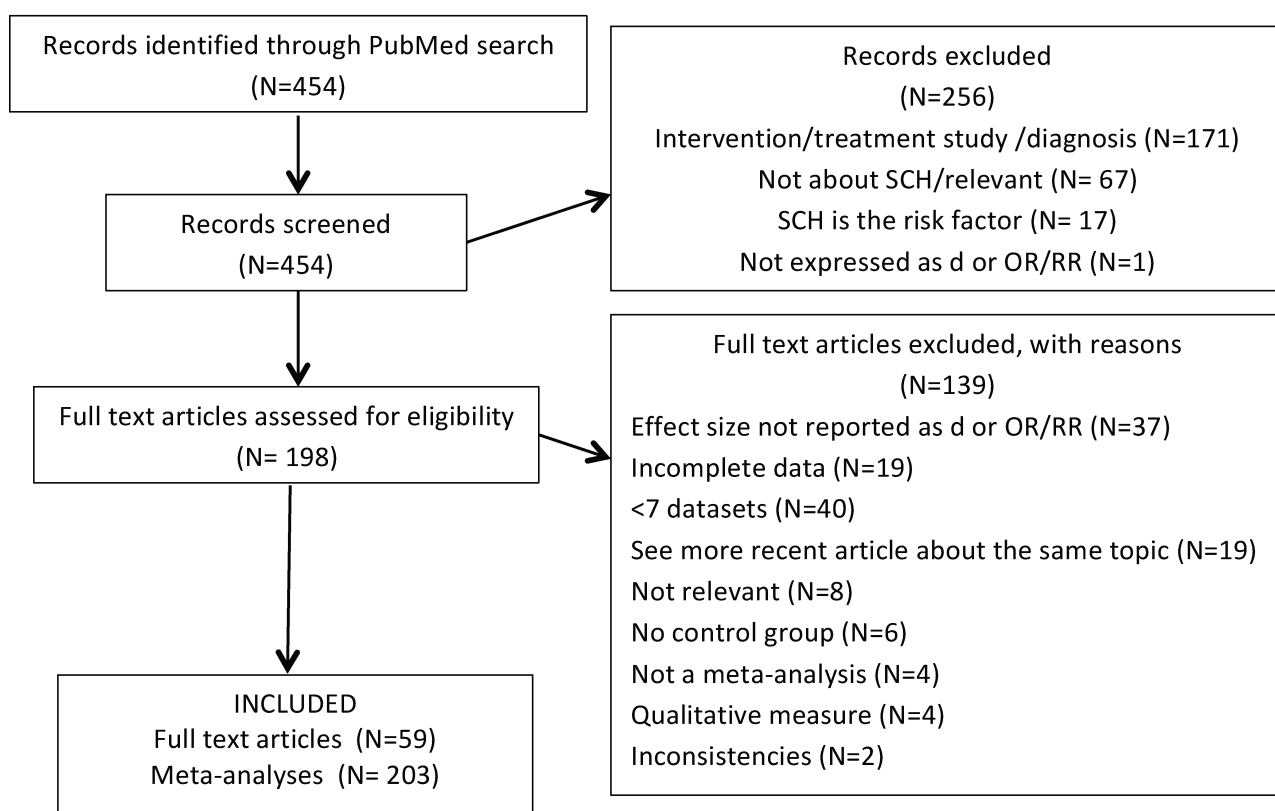


Fig S5. Flow diagram Alzheimer's disease (AD)

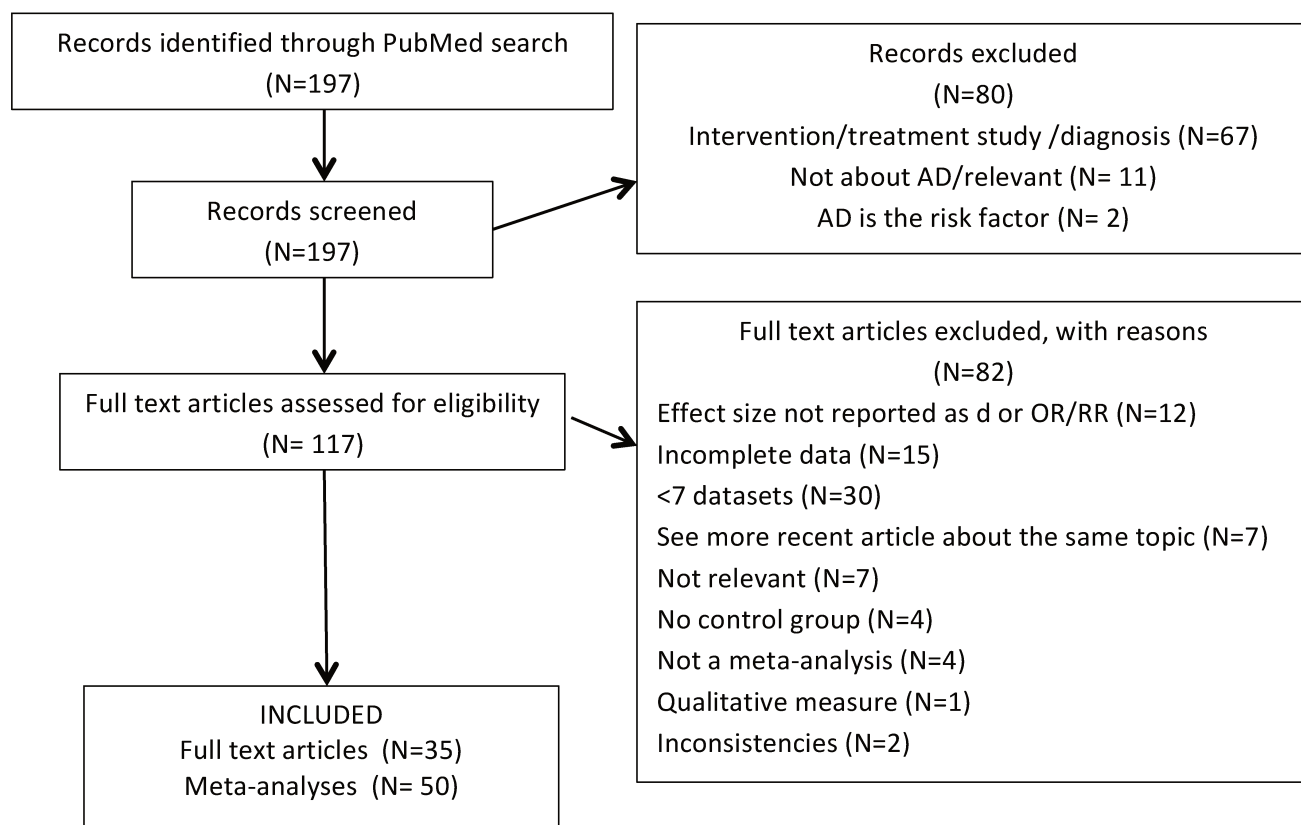


Fig S6. Flow diagram Epilepsy (Epi)

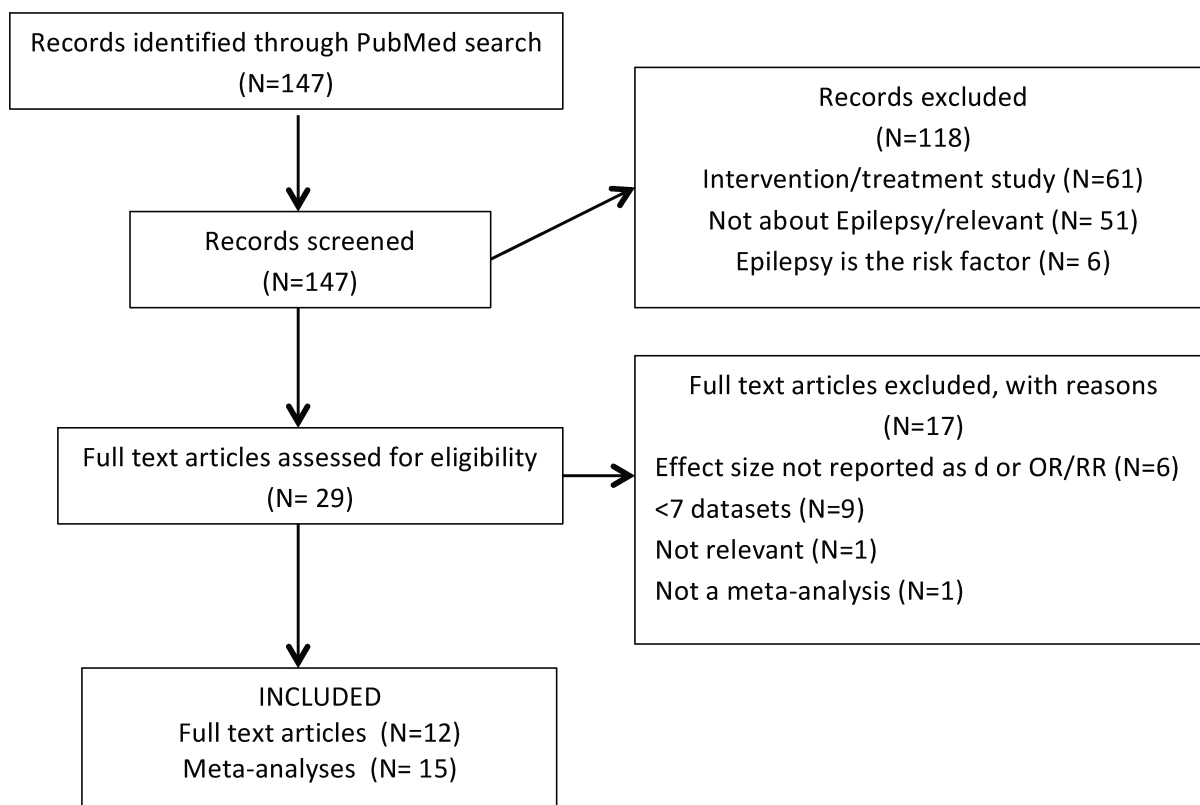


Fig S7. Flow diagram Multiple Sclerosis (MS)

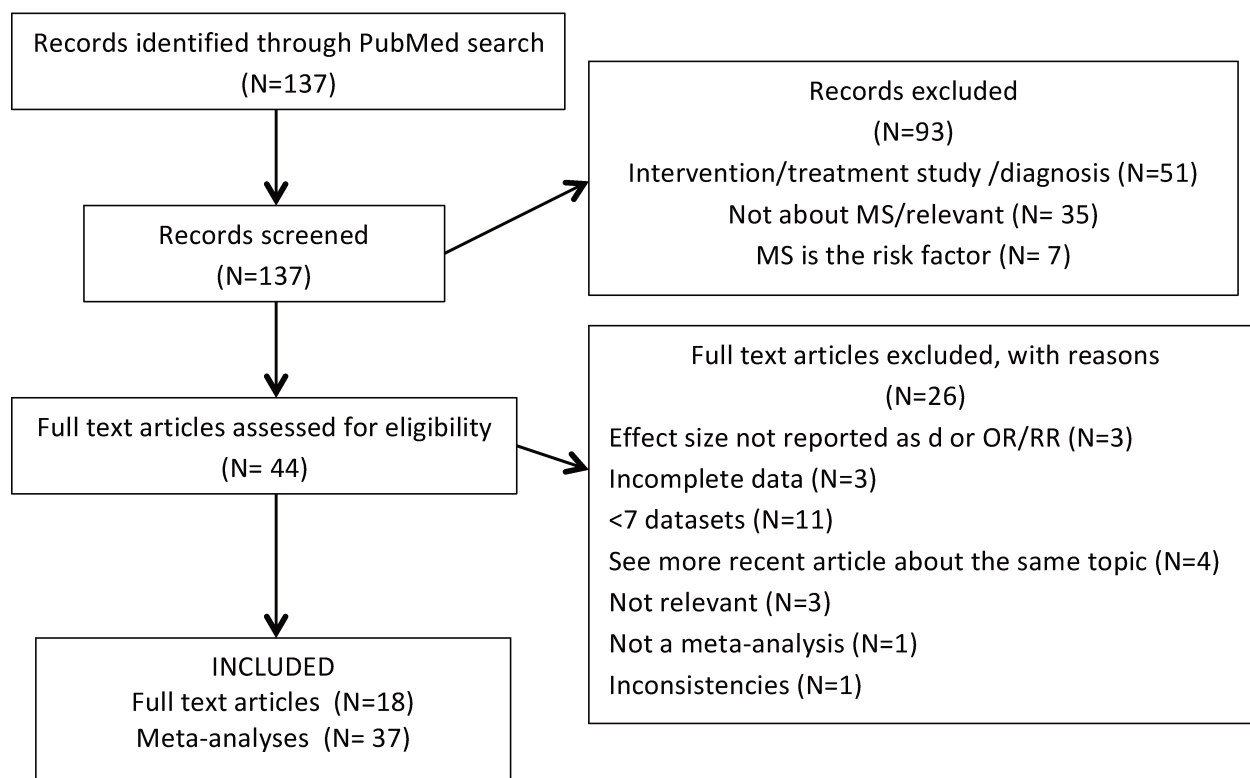


Fig S8. Flow diagram Parkinson's Disease (PD)

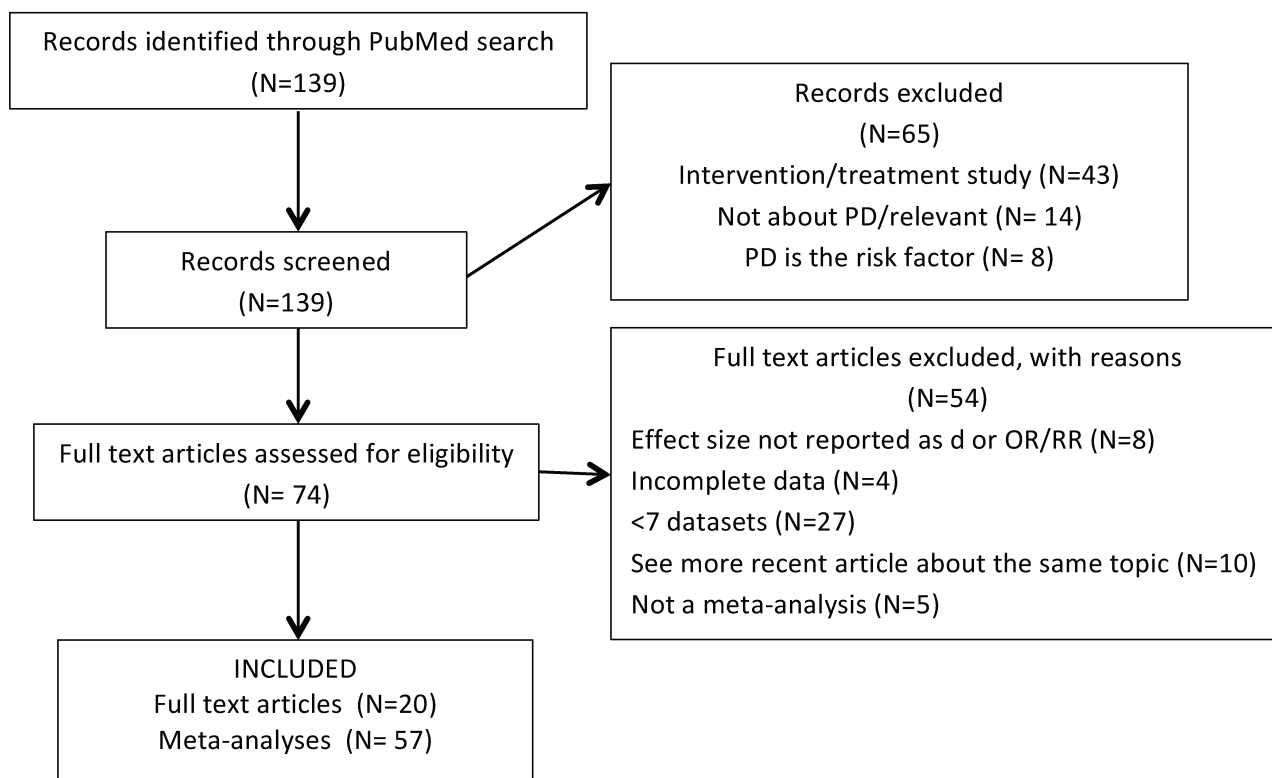


Fig S9. Flow diagram Breast Cancer (BC)

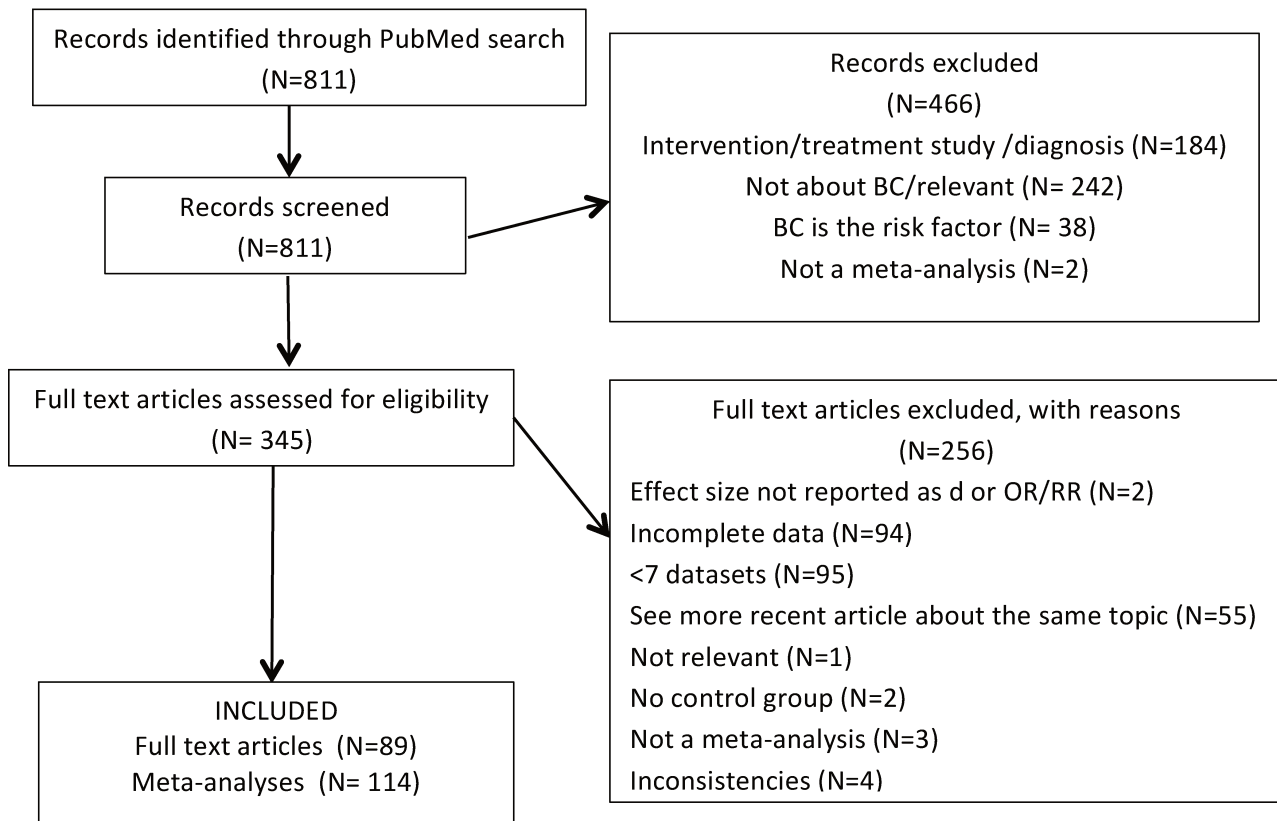


Fig S10. Flow diagram Glaucoma (GLAU)

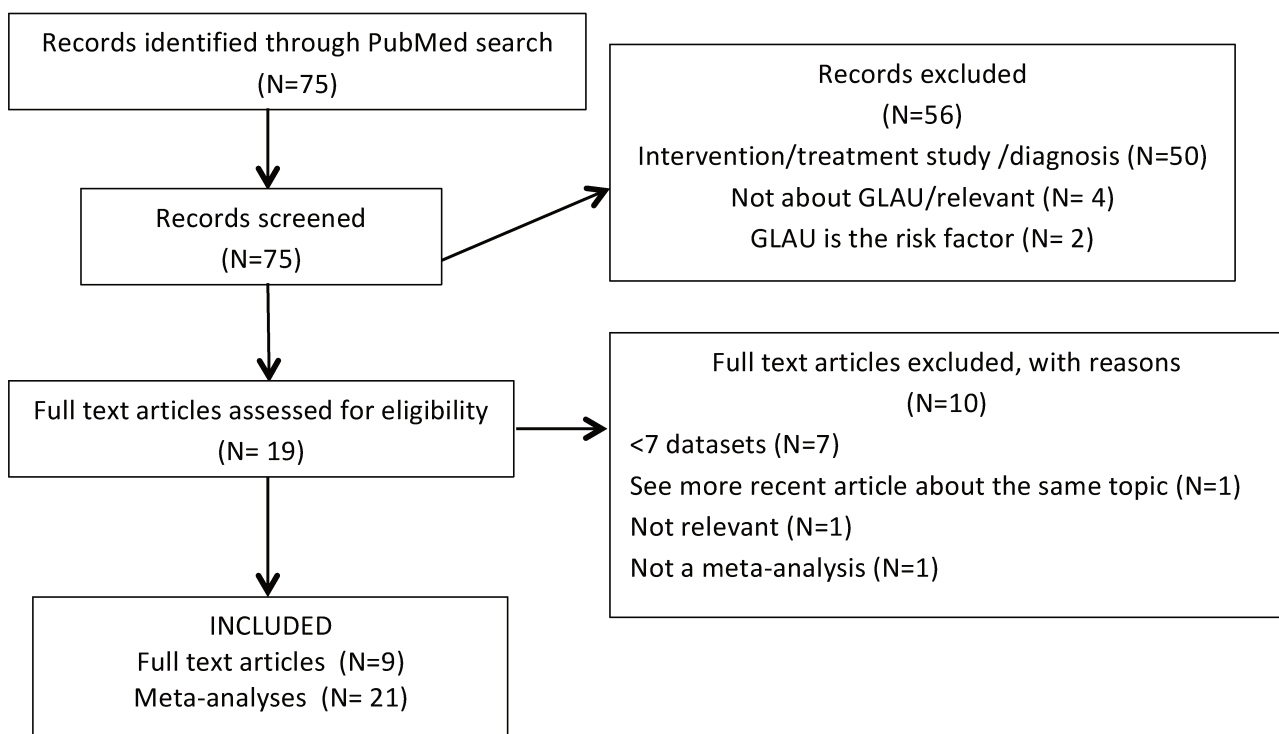


Fig S11. Flow diagram Psoriasis (PSO)

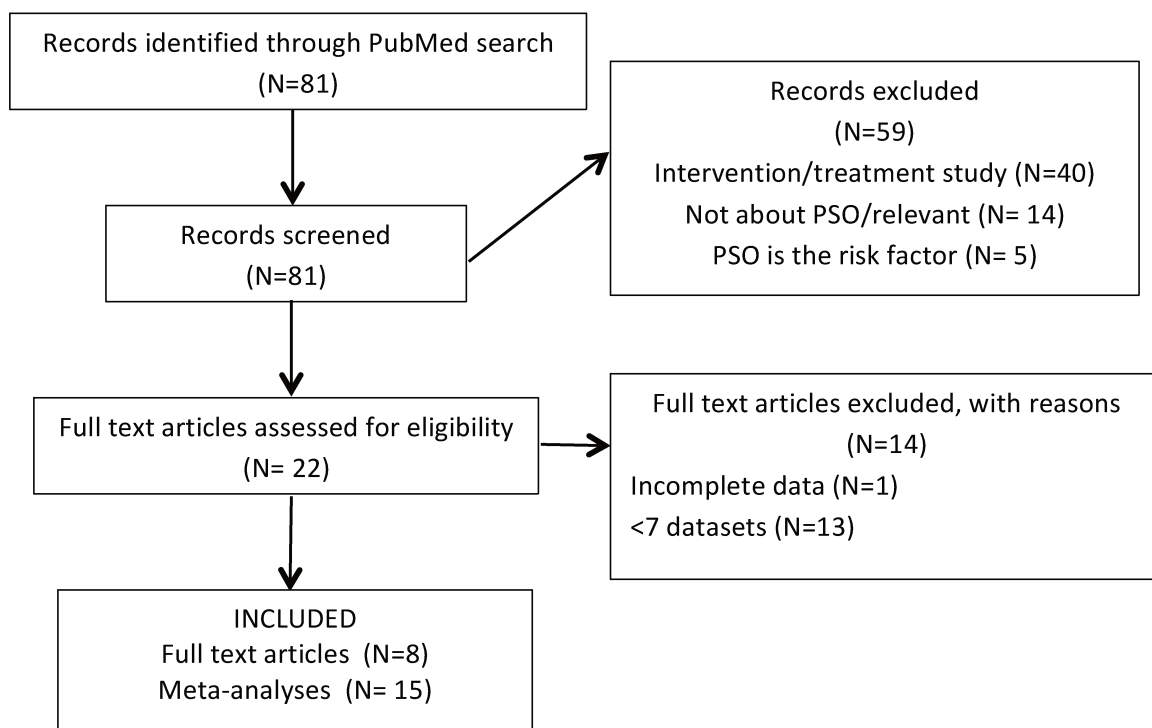
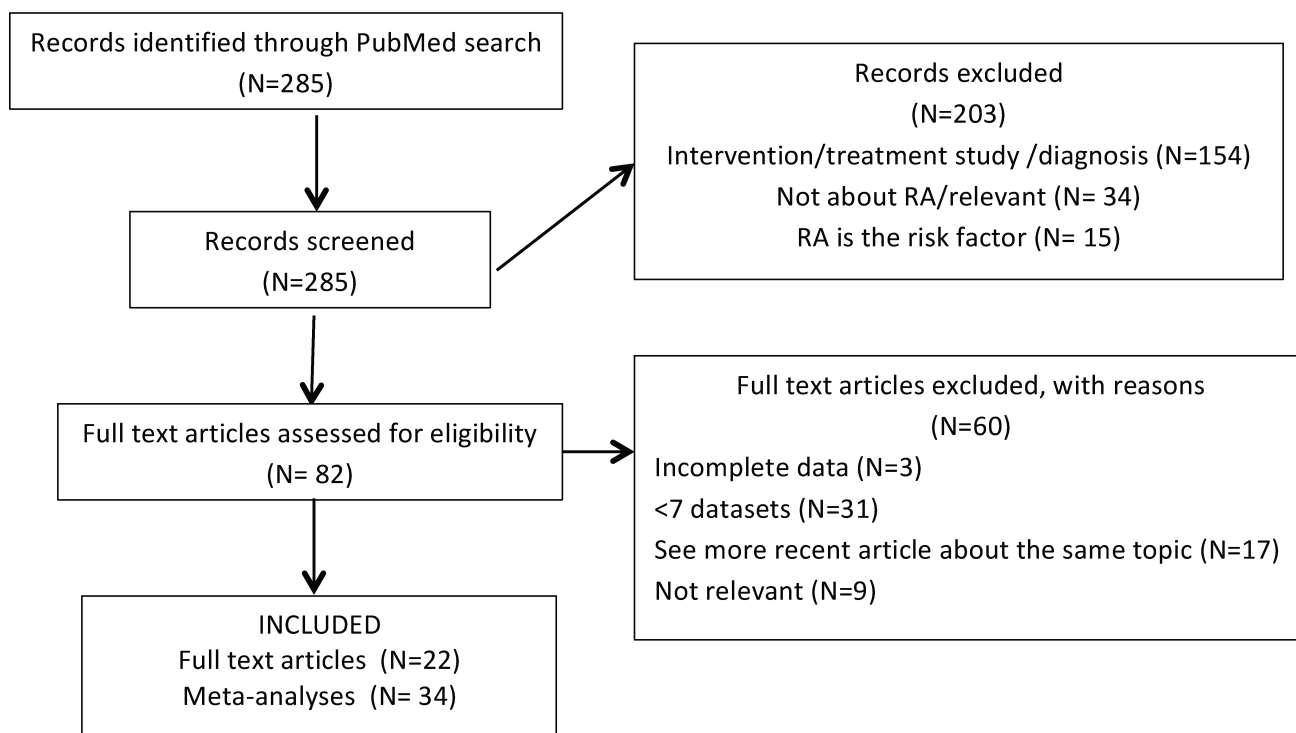


Fig S12. Flow diagram Rheumatoid Arthritis (RA)



ADHD																																				
Marker	# of lines	ID	Year	Initial study	IF	ES	CI low	CI high	p	# pat	# HC	[LSES]	Largest study	Year	Author	IF	ES	CI low	CI high	p	# pat	# HC	[LSES]	Meta-analysis		# pat	# HC	[LSES]	p							
																								Year	Author					IF	ES	CI low	CI high	p		
Urinary HVA	1	O	1971	Wender	N	14.7	0.29	-0.75	1.33	9	6	0.526	2002	Kutaba	N	3.1	0.5	-0.09	1.04	37	21	0.909	2012	Scassellati	N	9	0.11	-0.37	0.15	141	106	0.1992	15.6667			
Stim onset latency	1	O	1981	Babity	N	5.1	0.55	-0.303	1.4	11	11	0.9979	2004	Heurip	N	3.7	0.39	0.83	0.94	88	27	0.1074	2009	Corlese	N	7	-0.13	-0.35	0.09	219	148	0.2370	13.287			
SDSN-waldmet rs605 allelic	1	O	2005	Fried	N	3.5	1.01	0.5995	1.7325	83	96	0.0095	2006	Brookes	N	14.9	0.99	0.7955	1.2989	776	776	0.0101	2009	Foreo	N	14	0.68	0.89	1.13	3505	3508	0.0202	274.795			
ADRA2A MspI rs180544 allelic	1	O	2001	Xu	N	3.2	1.2	0.75	2	94	94	0.1632	2008	Brookes	N	14.9	1.06	0.87	1.15	776	776	0.0488	2009	Glzer	N	11	0.99	0.89	1.11	2274	2240	0.01005	206.777			
4-HTTLPR (5-HTT) rs10443G allelic	1	O	2001	Mavrou	N	3.2	1.12	0.8368	1.5147	98	98	0.1333	2008	Xu	N	3.2	1.15	0.9635	1.3398	1020	1020	0.1598	2009	Foreo	N	17	1.09	0.98	1.23	3500	4545	0.0818	724.708			
SLC6A4 VNTR (5-HTT S/T) rs10112 allelic	1	O	2002	Zengou	N	2.4	1.3	0.825	2.056	71	128	0.26236	2008	Xu	N	3.2	1.04	0.8702	1.263	1020	1020	0.0392	2009	Foreo	N	10	0.97	0.87	1.1	2533	2561	0.03046	253.3			
REM sleep %	1	O	1982	Greenhill	Y	5.1	-1.27			0.01	9	11	2.3093	2004	Huang	N	3	0.13	-0.31	0.56	88	27	0.2358	2009	Corlese	N	7	0.17	-0.05	0.39	219	148	0.38835	13.2873		
4-HTTR gene rs256 allelic	1	O	2002	Shaw	Y	14.9	1.49			0.0007	273	273	0.38978	2006	Brookes	N	14.9	0.87	0.6576	1.119	776	776	0.1923	2009	Foreo	N	10	1.09	0.96	1.19	3028	2952	0.0818	302.3		
DRD4 promoter insertion/deletion	1	O	2001	Barr	N	2.3	1.35	0.85	2.15	82	82	0.3001	2004	Kustanovich	Y	14.9	0.71			0.0007	535	535	0.3425	2012	Wu	N	9	1.09	0.91	1.3	1825	1824	0.0818	202.778		
Event related potential adult (Pz)	1	O	2003	Wiersma	Y	3	0.64			0.0431	23	19	1.10083	2010	Herrmann	Y	3.6	0.95			0.0002	34	34	1.7231	2013	Geburek	N	7	0.42	-0.003	0.84	106	161	0.2618	23.7145	
COMT valmet rs10443 allelic	1	O	1999	Eisenberg	Y	3.2	0.89			0.004	48	48	0.65998	2006	Brookes	N	14.9	1.02	0.89	1.12	776	776	0.0191	2009	Foreo	N	15	0.99	0.91	1.08	3242	3109	0.01005	292.625		
SNRP2 rs3745544 allelic	1	O	2009	Barr	N	14.9	1.04	0.5906	1.8421	87	87	0.00922	2003	Kustanovich	N	14.9	1.24	0.7955	1.8473	364	364	0.2151	2009	Foreo	Y	7	1.19			0.028	1435	1099	0.19978	209		
Stim efficiency (cognitiv test)	1	O	1991	Babity	N	5.1	-0.21	-1.05	0.63	11	11	0.3059	2004	Huang	N	3	0.39	-0.04	0.83	88	27	0.1074	2009	Corlese	Y	7	0.25			0.008	219	148	0.34545	13.2873		
DRD2-3'UTR rs180497 allelic	1	O	1991	Comings	Y	30	5.7			0.0001	104	108	1.54756	2004	Kustanovich	N	14.9	1.08	0.82	1.41	535	535	0.077	2012	Wu	Y	7	1.05			0.029	1893	1845	0.50078	210.333	
DAT 10 R 3'UTR VNTR allelic	1	O	1995	Cook	Y	11.2	2.84			0.0073	56	56	1.0438	2004	Kustanovich	N	14.9	0.92	0.75	1.2	535	535	0.0834	2009	Glzer	Y	34	1.1			0.028	4832	4818	0.09531	142.118	
Right caudate size (cm3)	1	I	1994	Castellanos	Y	14.7	0.45			0.0275	59	48	0.91621	2007	Castellanos	Y	13.8	0.43			0.0329	80	50	0.7795	2012	Frost	Y	7	0.56			0.0269	218	227	1.01373	13.2425
DAI imaging	1	I	1999	Dougherty	Y	39	2.37			0.0001	6	30	4.26971	2009	Volkov	Y	30	0.62			0.003	53	44	1.1246	2012	Fisnar-Pok	Y	9	0.23			0.04	169	173	0.41917	18.7778
Event related potential adult (Nz)	1	O	2003	Wiersma	N	3	0.41	-0.21	1.02	23	19	0.74366	2010	Herrmann	N	3.6	0.28	-0.2	0.76	34	34	0.9079	2013	Geburek	Y	7	0.5			0.0001	166	161	0.5069	23.7145		
DRD4 148-bp carrier	1	O	1999	Daly	Y	14.9	2.52			0.0001	118	111	0.92426	2004	Kustanovich	N	14.9	1.24	1	1.54	535	535	0.2151	2012	Wu	Y	12	1.26			0.0086	2262	2248	0.23111	190.167	
DRD4 gene T/C carrier	1	O	1998	Lalonde	Y	14.9	3.01			0.011	39	39	1.01964	2004	Kustanovich	N	14.9	1.21	0.94	1.56	535	535	0.1926	2010	Smith	Y	28	1.33			0.0001	3034	4210	0.2618	198.323	
Preschool interference control	1	O	1998	Hughes	Y	5.4	0.68			0.02	40	40	1.062	2007	Von Stauffenberg	N	17	0.04	-0.04	0.12	133	643	0.0728	2011	Pauli-Pott	Y	9	0.55			0.001	562	952	0.97959	62.4444	
Working memory, visuospatial	1	O	1989	Gorenstein	Y	1.7	0.667			0.0242	21	26	1.2098	2009	Mahone	N	7	0.09	-0.26	0.45	60	60	0.1632	2012	Kasper	Y	29	0.74			0.0001	830	1040	1.34221	28.6207	
Serum ferritin	1	O	2004	Konradi	Y	4.3	-1.26			0.0001	33	27	2.26839	2012	Derfrancesco	N	2.2	-0.05	-0.26	0.216	101	98	0.0607	2012	Carlson b	Y	8	-0.28			0.0004	221	1814	0.50786	40.121	
Theta/beta ratio all	1	O	1999	Monstra	Y	3.6	1.45			0.0001	304	64	2.63001	2013	Loe	N	2.2	0.23	-0.009	0.47	352	83	0.4172	2013	Arns	Y	9	0.45			0.0001	1253	917	0.81821	193.222	
Urinary MHPG	1	O	1971	Wender	N	14.7	0.32	-1.36	0.72	9	6	0.58042	1984	Shan	Y	9.2	-0.36			0.0438	73	57	0.653	2012	Scassellati	Y	15	0.41			0.0001	298	236	0.74366	19.8667	
Gender effect on CTP test (commission er)	1	O	1989	Horn	N	3.1	0.39	-0.48	0.69	54	0	0.16324	1998	Arns	Y	1.7	0.27			0.01	360	0	0.4897	2012	Hasson	Y	8	0.31			0.0001	1097	0	0.54228	137.123	
Urinary NE	1	O	1970	Sapozot	N	39	0.23	-0.71	1.17	16	6	0.41717	2007	Pevetkova	Y	1.9	0.76			0.0076	99	17	1.0004	2012	Scassellati	Y	7	0.41			0.0001	197	91	0.74366	24.436	
Urinary monoamine	1	O	1992	Shue	N	2.8	0.4	-1.02	0.22	22	18	0.78252	2009	Coulinho	Y	19	0.417			0.002	186	80	0.7664	2012	Kasper	Y	34	0.69			0.0001	1888	1086	1.21512	29.0588	
Saltatory cortisol	1	O	1999	Jansen	N	4.6	0.06	-0.74	0.86	10	15	0.10883	2009	Van West	Y	3.7	-0.66			0.0049	75	25	1.1971	2012	Scassellati	Y	8	-0.31			0.0001	322	300	0.54228	40.25	
Strategy scores CANTAB	1	O	1999	Kempton	N	5.6	0.35	-1.068	0.374	30	15	0.63483	2010	Gea	Y	5.4	-0.45			0.0001	279	173	0.843	2011	Chamberlain	Y	11	0.51			0.0001	650	453	0.93248	93.0605	
Go/NoGo commission error (EOC)	1	O	2006	Borger (F-ast)	N	3.4	-0.101	-0.665	0.462	27	22	0.18319	2010	Uebel (Fast)	Y	5.4	0.653			0.0001	205	53	1.1844	2012	Metin	Y	20	0.44			0.0001	1300	2207	0.78007	47.913	
Regulatory problem bat 2	1	O	1989	Zuckerman	Y	5.1	0.88			0.01	23	33	1.77752	2001	Mston	Y	0.1	0.21			0.001	373	1069	0.3809	2011	Hemm	Y	13	0.36			0.0025	849	1137	0.65297	65.8077
Time off-task in classroom	1	O	1992	Werry	Y	2.8	1.867			0.0015	11	10	2.8765	2002	Abekutt	Y	3.1	0.38			0.0001	602	902	0.0962	2008	Keller	Y	23	0.73			0.0001	1176	1176	1.32027	53.3304
Preschool working memory	1	O	1997	Mariani	Y	2.9	0.72			0.001	34	30	1.30994	2002	Kaif	Y	3.1	0.36			0.03	108	133	0.653	2011	Pauli-Pott	Y	10	0.52			0.001	569	794	0.58042	56.9
Colorings (parents)	1	O	1978	Williams	Y	5.1	0.52			0.01	26	0	0.94318	2004	Bateman	Y	3	0.12			0.0454	277	0	0.2177	2012	Naga	Y	20	0.18			0.0007	794	0	0.34548	39.3
Emos CANTAB test	1	O	1999	Keupman	Y	5.6	-0.29			0.001	39	15	2.3098	2004	Rhodes	Y	4.1	-0.76			0.0001	75	70	1.3024	2011	Chamberlain	Y	8	-0.68			0.0001	352	261	1.50514	44
Preschool response inhibition	1	O	2002	Sonuga-Barke	Y	2.9	0.63			0.04	48	113	1.14269																							

Autism		Initial study										Largest study										Meta-analysis															
Marker	# of lines	ES	Type	Year	Author	Sign	IF	ES	CI low	CI high	p	# obs	# HC	[ES(ES)]	Year	Author	Sign	IF	ES	CI low	CI high	p	# obs	# HC	[ES(ES)]	Year	Author	Sign	IF	ES	CI low	CI high	p	# obs	# HC	[ES(ES)]	
Verbal IQ (adult)	g	I	1988	Courchesne	N	51.7	0.05	-0.78	0.68	18	12	0.0907	1994	Courchesne	N	2.9	-0.06	-0.44	0.329	50	53	0.126828	2008	Starfield	N	7	0.07	-0.169	0.309	131	149	0.12693	18,174,286				
ST192 VN19 I2 (other) TD	OR	G	1997	Cook	N	14.9	1.65	0.98	3.19	86	172	0.5508	2004	Courtho	N	14.9	1.58	0.95	2.439	182	364	0.444686	2008	Huang	N	8	-1.13	0.819	1.568	809	1618	0.12222	1,011,125				
NAA global grey matter adults	g	I	2002	Murphy	N	13.8	0.75	-0.03	1.52	12	18	1.3504	2009	Kiemane	N	3.7	0.41	-0.22	1.05	29	19	0.743953	2012	Ipsier	N	7	-0.01	-0.498	0.398	190	192	0.01611	14,285,145				
HXB1 nINSINS insertion	OR	G	2000	Ingram	N	3.1	1.45	0.85	2.5	50	50	0.3716	2010	Muscarella	N	5.6	1.08	0.84	1.38	247	247	0.076961	2011	Sang	N	8	1.14	0.97	1.33	683	784	0.13103	85,375				
Neonatal jaundice	OR	O	1979	Finegan	N	5.4	0.9	0.39	164	23	23	2.0794	2009	Budtmeyer	N	5.1	1.29	0.99	1.07	1098	5681	0.254842	2011	Amin	YY	13	1.43			0.0001	2672	5690	0.35767	220,923,038			
Serum calcium	g	I	1987	Gaffney	N	6.6	0.27	-0.37	0.9	13	35	0.4897	1999	Epas	N	7.7	0.53	-0.07	0.72	51	51	0.989524	2008	Fraser	YY	10	0.48			0.0001	253	259	0.47962	75.7			
NAA frontal children	g	I	2001	Hisaka	N	2.7	0.153	-0.23	0.534	55	51	0.2775	2010	Mori	N	1.7	-0.4	-0.92	0.12	70	18	0.72552	2012	Aoki	YY	8	-0.35			0.0001	764	531	0.63483	55.5			
Paternal age 30-39 / s29	RR	O	2004	Glasson	N	13.8	1.2	0.993	1.6957	403	1208	0.1563	2009	Grether	Y	4.8	1.67			0.0001	17711	5629545	0.512824	2011	Hultman	Y	12	1.22			0.0009	21837	7878495	0.19883	1819,75		
Methionine plasma/serum	g	O	1996	DiEgidio	N	2.1	0.12	-0.54	0.305	49	46	0.2177	2006	James	Y	3.2	-1.08			0.0001	89	73	2.283381	2012	Frustao	Y	7	-0.8			0.043	345	324	1.45104	45,285,145		
Verbal IQ (adult)	g	I	1992	Kleiman	N	8.2	0.35	-0.44	1.136	10	17	0.6348	2000	Garper	Y	9.9	-0.66			0.0228	42	29	1.015728	2008	Starfield	YY	8	-0.43			0.0003	183	132	0.77993	22,875		
Paternal age 50 and older / s29	RR	O	2004	Glasson	N	13.8	1.1	0.443	2.9624	141	450	0.0953	2009	Grether	Y	4.8	2.45			0.0001	1813	3256579	0.896088	2011	Hultman	YY	7	2.46			0.0001	8410	3869744	0.00016	1201,4286		
RIT1LPS S41 family-based trio	OR	G	1997	Cook	Y	14.8	1.86			0.0316	86	112	0.5668	2004	Courtho	N	14.8	1.11	0.82	1.51	182	364	0.10438	2008	Huang	N	13	1.03	0.842	1.265	1031	2692	0.03995	75,307,691			
HXA1 A218G rs1091154 allelic	OR	G	2000	Ingram	Y	3.1	1.88			0.0418	50	50	0.6313	2002	Devlin	N	2.3	0.83	0.59	1.16	221	221	0.18633	2011	Sang	N	14	0.98	0.76	1.25	1076	1214	0.0001	76,857,143			
NAA global grey matter children	g	I	1999	Otsuka	Y	2.7	0.81			0.0458	21	10	1.4892	2001	Hisaka	N	2.7	-0.16	-0.54	0.23	55	51	0.290208	2012	Ipsier	N	9	-0.28	-0.594	0.036	237	179	0.50786	26,333,333			
Homocysteine plasma/serum	g	O	2004	James	Y	6.5	0.49			0.01	20	33	0.8868	2006	James	N	3.2	-0.24	-0.597	0.079	80	73	0.435313	2012	Frustao	N	7	-0.09	-0.581	0.389	238	218	0.16324	24,142,857			
NAA frontal adults	g	I	2002	Murphy	Y	13.8	0.99			0.0084	11	18	1.977	2007	One	N	3.2	0.51	-0.18	1.19	14	21	0.929333	2015	Aoki	N	7	0.12	-0.31	0.54	90	113	0.17362	12,857,143			
Total Brain Volume (adults)	g	I	1995	Phven	Y	14.7	0.77			0.0162	22	20	1.3968	2003	Hardan	N	1.4	0.27	-0.166	0.706	40	41	0.489728	2008	Starfield	Y	8	0.27			0.013	150	181	0.48973	18,75		
Verbal IQ (adult)	g	I	1988	Courchesne	Y	51.7	1.22			0.0026	18	12	2.2128	1994	Courchesne	Y	2.9	1.12			0.0001	43	53	2.01445	2008	Starfield	N	8	-0.12	-0.36	-0.13	120	187	0.17176	16.24		
Motor coordination	g	O	1981	Wienasy	Y	7.7	0.9			0.011	21	15	1.6324	2008	Gombaszch	Y	5.4	0.6			0.0003	115	44	1.58828	2010	Fourmer	YY	51	1.2			0.0001	957	1087	2.17651	18,764,705	
Paternal age 40-49 / s29	RR	O	2004	Glasson	Y	13.8	1.88			0.0171	190	551	0.4574	2009	Grether	Y	4.8	2.23			0.0001	10047	3738543	0.892002	2011	Hultman	YY	12	1.78			0.0001	12405	4764773	0.17661	1033,75	
Maternal age 340 or > 35 / s29	RR	O	2005	Larson	Y	4.8	3.19			0.0012	270	7339	0.7339	2009	Grether	Y	4.8	1.84			0.0001	14007	3002371	0.697964	2012	Sander	YY	10	1.52			0.0003	8619	3193856	0.41871	861,9	
Total Brain Volume (children)	g	I	2001	Courchesne	Y	8.2	1.42			0.0002	30	12	2.5796	2002	Sparks	Y	8.2	1.03			0.0003	29	26	1.888214	2008	Starfield	YY	9	0.41			0.0004	160	188	0.43664	17,777,779	
Sensory symptoms	g	O	1998	Ermer	Y	1.7	2.11			0.0001	38	1075	3.8271	2007	Tomcheck	Y	1.7	1.71			0.0001	281	281	3.101598	2008	Ben-Sasson	YY	15	1.5			0.0001	1190	2184	2.7207	75,333,333	
Minor Physical Anomalities	g	O	1975	Steg	Y	3.7	0.68			0.0119	28	31	1.2234	1977	Walker	Y	3.7	0.97			0.0001	74	74	1.789385	2010	Ozgen	Y	7	0.84			0.0001	320	382	1.94392	47,142,857	

Data collected for Autism: Data for the initial study, largest study and meta-analysis. Column C: type of study, O is for "Other", G for "Genetic" and I for "Imaging"
 First author and year of indicated in bold correspond to initial studies that were not given by the corresponding meta-analysis

Schizophrenia

Table with columns: Marker, Chr, Position, P-value, R-squared, and other genomic data. The table lists various genetic markers and their associations with schizophrenia across different chromosomes and populations.

Data collected for schizophrenia. Data for the initial study, largest study and meta-analysis. Column C is for "Other", G for "Genetic", H for "Hereditary", and I for "Cognitive or behavioral".

First author and year of publication is bolded to correspond to initial studies that were not given the corresponding meta-analysis.

Alzheimer																																								
Marker	# of times	ES	type	Year	Author	Initial study	Sign	IF	ES	CI low	CI high	p	# pat	# HC	LFC3	Largest study	Year	Author	Sign	IF	ES	CI low	CI high	p	# pat	# HC	LFC3	Meta-analysis	Year	Author	Sign	# stud	ES	CI low	CI high	p	# pat	# HC	LFC3	# meta
ABCA1 rs3811	1	0.003	G	2004	Kozlov (set A)	N	5.2	1.29	0.8355	1.8852	394	176	0.2545	2004	Li	N	2	1.62	0.76	1.38	418	376	0.10198	2012	Chang	N	8	1.1	0.96	1.28	2118	1852	0.05311	254.75						
ABCA1 rs3811	1	0.003	G	2004	Kozlov (set B)	N	5.2	1.29	0.8355	1.8852	394	176	0.2545	2004	Li	N	2	1.62	0.76	1.38	418	376	0.10198	2012	Chang	N	8	1.1	0.96	1.28	2118	1852	0.05311	254.75						
ABCA1 rs3811	1	0.003	G	2004	Kozlov (set C)	N	5.2	1.29	0.8355	1.8852	394	176	0.2545	2004	Li	N	2	1.62	0.76	1.38	418	376	0.10198	2012	Chang	N	8	1.1	0.96	1.28	2118	1852	0.05311	254.75						
ABCA1 rs3811	1	0.003	G	2004	Kozlov (set D)	N	5.2	1.29	0.8355	1.8852	394	176	0.2545	2004	Li	N	2	1.62	0.76	1.38	418	376	0.10198	2012	Chang	N	8	1.1	0.96	1.28	2118	1852	0.05311	254.75						

Data collected for Alzheimer's disease. Data for the initial study, largest study and meta-analysis. Column C: type of study. D is for "Other", G for "Genetic", I for "Imaging" and C for "Cognitive or behavioural test". First author and year of indicated in bold correspond to initial studies that were not given by the corresponding meta-analysis.

Epilepsy

Marker	# of lines	Initial study						Largest study						Meta-analysis																						
		ES	Year	Author	Sign	IF	ES	CI low	CI high	p	# pati	# HC	[I-FES]	Year	Author	Sign	IF	ES	CI low	CI high	p	# pati	# HC	[I-FES]	#pati/HC											
rs13111 T/T versus C/C homozygote	OR	G	2009	Kanavabo	N	31.2	1.96	0.912	4.2141	54	54	1,2684	2005	Cavalletti	N	3.9	0.94	0.6139	1.4408	333	202	0.0615	2010	Kaufman b	N	13	1.2	0.928	1.562	1016	1262	0.18232	78,1538			
GABRG2 C588T rs211037 T/C allelic	OR	G	2002	Kanavabo	N	7.7	1.10	0.77	1.74	135	154	0.1398	2006	Kinirons	N	2.2	1.1	0.87	1.4	699	283	0.0953	2013	Haeman	N	8	0.79	0.56	1.11	1871	1387	0.23572	233,875			
ABO1 C345T C/C vs T/T homozygote	OR	G	2003	Siddiqui	N	51.7	1.27	0.74	2.18	315	200	0.239	2007	Kwan	N	1.8	0.9	0.51	1.58	746	179	0.1054	2010	Nurmohamed	N	9	1.07	0.76	1.51	2454	1542	0.06766	272,667			
Heart rate variability: sympathetic activity Low Fr	d	O	1997	Massimini	Y	3.5	-0.41			0.0001	30	20	2.6576	2008	Hallgrimsdottir	N	2.2	0.01	-0.3	0.32	78	83	0.0181	2012	Lotito	N	18	-0.18	-0.71	0.38	458	594	0.32848	27,1311		
Age of onset AcdE4 carrier	d	O	2000	Breitmann	Y	0.2	-1.08			0.0045	43	43	1.9589	2005	Cavalletti	N	9.9	0.28	0.87	0.12	181	181	0.5079	2010	Kaufman	Y	8	0.74			0.007	728	728	1.34221	93	
Alcohol consumption <50g	RR	O	1988	Hg	N	51.7	1.1	0.5	2.4	145	128	0.0953	1997	Leone	Y	8.2	1.81			0.0332	152	296	0.5933	2010	Samokhvalov	Y	7	1.29			0.0254	698	1162	0.25464	99,7143	
LS in plasma	d	O	1998	Pettila	N	2.2	0.57	-0.099	1.239	15	22	1.6039	2009	Lindstrom	N	3.6	0.27	-0.083	0.592	91	63	0.4897	2012	Tu	YY	8	0.7			0.0001	261	964	0.24966	24,625		
Resting motor threshold	d	O	1992	Reiners	Y	38	-1.27			0.0002	11	71	2.3035	2010	Badway	N	11.2	-0.36	-0.797	0.081	65	32	0.663	2012	Brigo	YY	16	-0.3			0.0005	265	424	0.54414	16,5425	
Heart rate variability: vagal activity High Fr	d	O	1997	Massimini	Y	3.9	-0.81			0.0069	30	20	1.4692	2008	Hallgrimsdottir	N	2.2	-0.24	-0.55	0.07	78	83	0.4353	2012	Lotito	YY	18	0.69			0.0002	488	864	1.25152	27,1311	
Autism in first 24 hr risk	RR	O	1976	Stark	N	3.7	0.18	0.0283	1.2913	16	48	1.7148	1983	Jacobson	Y	1.1	0.38			0.0001	337	653	0.6676	2008	Armet	YY	10	0.56			0.0001	370	1742	0.21262	43	
Fractional anisotropy uncinate fasciculus	d	I	2007	Rodrigo	Y	3.5	-1.26			0.0101	10	10	2.2854	2008	Diehl	Y	3.9	-0.61			0.0275	18	10	1.6506	2012	Otte	YY	12	-0.82			0.0001	156	158	1.48792	13
Cytomegalovirus infection in Africa	OR	O	1989	Mignard	Y	0.5	2.5			0.0001	242	385	0.9163	2003	Nsengyuma	Y	3.9	3.2			0.0001	324	648	1.1632	2010	Quet	YY	11	3.37			0.0001	1658	5628	1.21491	150,727
Alcohol consumption case-control	RR	O	1987	Ogumny	Y	3.9	2.4			0.003	156	156	0.8755	1987	Ogumny	Y	3.9	2.4			0.003	156	156	0.8755	2010	Samokhvalov	YY	9	2.18			0.0001	634	1288	0.2819	103,778
Autism male/female as risk	RR	O	1981	Tea	Y	3.7	0.4			0.0038	23	79	0.9163	1991	Tuchman	Y	5.1	0.45			0.0042	43	352	0.7985	2008	Armet	YY	14	0.55			0.0001	337	1193	0.55786	24,0714
Toxocara infection	OR	O	1979	Glickman	Y	4	2.58			0.018	84	108	0.9478	2008	Niccolini	Y	3.9	2.85			0.0018	231	201	1.0473	2012	Quattrocchi	YY	7	1.62			0.0001	850	1017	0.65233	121,429

Data collected for Epilepsy: Data for the initial study, largest study and meta-analysis. Column C: type of study. O is for "Other", G for "Genetic and I for "Imaging". First author and year of indicated in bold correspond to initial studies that were not given by the corresponding meta-analysis

Parkinson

Marker	# of genes	ES	Year	Initial study				# of cases	# HC	[L-R]	Largest study				# of cases	# HC	[L-R]	Meta-analysis				# of cases	# HC	[L-R]	P-value										
				Year	Year	Year	Year				Year	Year	Year	Year				Year	Year	Year	Year					Year	Year	Year	Year	Year	Year				
MAP1B/STH14H caucasian allelic	OR	G	2001	Hasegawa	N	112	137	0.97	1.93	187	297	0.91461	2011	PDGC USA rep	Y	0.5	0.76	0.0001	2807	2215	0.274437	2012	Li	YY	37	0.78	0.0001	22984	27987	0.248641	621,189				
L118_G1117 rs18944 C/T allelic	OR	G	2009	Nishimura	N	2	1.92	0.64	1.33	121	112	0.96338	2007	Wu	N	3	0.87	0.72	1.05	493	388	0.139262	2012	Zhu	N	8	0.99	0.768	1.256	1920	1709	0.01005	240		
Non 10 CAG rep in POKI3 allelic	OR	G	2003	Hannan	N	2	0.98	0.31	2.48	22	31	0.12783	2010	Santa	Y	2	1.56	0.004	641	202	0.041911	2012	Blalahan	YY	7	1.27	0.0004	1597	1884	0.278917	226,143				
L118_C1089 rs1020587 C/T allelic	OR	G	2009	Nishimura	N	2	1.05	0.59	1.88	122	112	0.94879	2008	Zhu	Y	2	0.73	0.004	533	530	0.345711	2012	Zhu	N	9	1	0.88	1.12	2117	2029	0.4	235,222			
Copper level in serum/plasma	d	O	1992	Limez	N	2.2	0.26	-0.19	0.7	39	39	0.47159	2006	Bocca	N	2.2	-0.43	0.94	0.08	91	18	0.779394	2013	Manara	N	13	0.16	-0.6	0.91	627	572	0.290208	48,338		
Fungicides exposure	RR	O	1992	Semchuk	N	8.2	1.63	0.81	3.29	130	260	0.48888	2009	Brigina	N	8.2	0.83	0.44	1.59	833	833	0.18633	2012	Van der Mark	N	9	0.99	0.71	1.4	1215	2890	0.01005	235		
PRKN rs628443 rs1801334 allelic	OR	G	2001	Oliver	N	8.2	1.37	0.44	4.25	158	100	0.91481	2003	Lincoln	N	4.6	0.69	0.37	1.3	265	187	0.371864	2012	Zhang	N	7	1.05	0.65	1.7	2215	1912	0.004569	171,286		
Dietary potassium	RR	O	1972	Kessler	N	4.8	0.55	0.3	1.1	228	228	0.54473	2008	Rapaglia	N	8.2	1.1	0.8	1.5	13955	8845	0.009531	2012	Noyce	N	13	0.91	0.72	1.15	20926	30343	0.094311	1540,38		
COMT rs105188Met (G/A) rs4690 allelic	OR	G	1996	Roca	N	2.4	1.01	0.74	1.39	139	173	0.00995	2008	Balaska	N	3.6	0.87	0.7	1.08	322	367	0.139262	2012	Wang	NN	24	0.98	0.92	1.05	3807	3942	0.020209	158,625		
POM1 rs456198 rs456198 allelic	OR	G	2009	Wang	N	4.6	1.07	0.52	2.3	160	160	0.06766	2007	Dick	N	3.2	0.84	0.827	1.05	736	919	0.061875	2012	Li	N	6	1	0.95	1.11	2682	997	0.002598	266,869		
PITx3 rs2281983 C/T allelic	OR	G	2009	Liu	N	8.2	1.01	0.92	1.25	669	669	0.06766	2009	Fuchs	N	8.2	1.07	0.92	1.25	669	669	0.067669	2012	Lang	NN	8	1.03	0.97	1.11	3472	3604	0.029359	434		
LINSO1 gene rs952480 A/G allelic	OR	G	2009	Saundersberger	N	4.6	1.05	0.72	1.53	348	342	0.04979	2010	Klebe	N	3.2	1.05	0.88	1.24	1798	1482	0.04879	2012	Agutierrez	NN	12	1.001	0.91	1.08	6777	6258	0.001	564,75		
Cancer antecedent cases/all	RR	O	1972	Kessler	N	4.8	0.71	0.2	2.26	228	228	0.34246	2008	Olson	N	8.7	1.04	0.96	1.12	6990	3200	0.009221	2012	Noyce	NN	7	1.01	0.94	1.09	9690	3413	0.00989	1384,71		
Well water exposure classic	RR	O	1989	Nanner	N	8.2	0.74	0.41	1.32	100	200	0.30111	2007	Dick-b (FD)	N	3.2	1.23	1.12	1.52	767	1989	0.207014	2012	Noyce	Y	26	1.21	0.012	5037	8144	0.13063	179,893			
Tea drinking case/control	RR	O	1981	Abadi	N	4.8	0.63	0.31	1.29	237	474	0.46204	2001	Paganelli-Hill	N	2.4	0.91	0.73	1.12	395	2320	0.094311	2012	Li	Y	8	0.85	0.02	4118	4250	0.162519	177,25			
APOE-ε2 carriers	OR	G	1994	Nander	N	8.2	0.76	0.19	3.25	79	44	0.27444	2009	Williams-Gray	N	3.6	1.03	0.72	1.47	505	478	0.029999	2009	Williams-Gray	Y	31	1.16	0.017	4198	10066	0.14842	135,419			
Alcohol consumption	OR	G	1999	Wu	N	8.2	0.7	0.2	7	35	105	0.35687	2004	Herman	N	8.6	1.08	0.59	1.28	1019	1023	0.078961	2012	Wang	YY	24	0.9	0.002	4320	21937	0.00563	208			
PARK1 rs47211 A/G caucasian allelic	OR	G	2009	Pankratz	N	8.2	0.97	0.83	1.14	807	887	0.03046	2009	Simon-Sanchez	N	35.2	0.98	0.88	1.04	3452	4765	0.045222	2012	Li	YY	12	0.91	0.0001	17528	18102	0.059311	1485,87			
Iron level in serum	d	O	1964	Cabrera	N	2.2	0.38	-0.01	0.71	81	80	0.63483	1997	Logrescino	Y	8.2	1.01	0.82	1.1	0.0001	104	382	1.831938	2013	Manara	NN	10	0.46	-0.98	0.08	0.0001	14402	19837	0.157004	885,81
GIA gene N2753 G/A caucasian allelic	OR	G	2004	Lavin	N	2.8	11.8	0.66	210.85	57	44	2.4681	2011	Do	Y	8.5	4.03	0.0001	3405	2924	1.398717	2012	Li	YY	15	3.51	0.0001	6908	3743	1.293508	933,867				
PARK1 rs23156 G/A caucasian allelic	OR	G	2009	Pankratz	N	4.2	0.98	0.82	1.17	807	887	0.03022	2011	Do	Y	8.5	0.827	0.0001	3426	2924	1.869951	2012	Li	YY	11	0.9	0.0003	14376	46291	0.105361	1306,91				
GIA gene N2753 G/A caucasian allelic	OR	G	2009	Pankratz	N	4.2	1.08	0.91	1.28	857	887	0.07096	2010	Hanna	Y	35.2	1.25	0.0001	2000	1986	0.223144	2012	Li	YY	21	1.17	0.0001	14402	19837	0.157004	885,81				
PARK1 rs1491842 G/C caucasian allelic	OR	G	2009	Pankratz	N	4.8	0.97	0.6	1.58	228	228	0.03046	2008	Raughens	Y	17.2	1.5	0.0001	13955	6845	0.459465	2012	Noyce	YY	19	1.98	0.0001	17139	7842	0.457426	902,053				
Head injury preceding all	RR	O	1971	Reynolds	N	39	0.13	0.01	2.54	36	43	0.24022	2005	Liss	Y	8.2	0.37	0.0001	501	713	1.121678	2012	Van der Werf	YY	10	2.13	0.0001	1890	2556	0.754212	185				
Hypertension preceding all	RR	O	1972	Kessler	N	4.8	1.06	0.78	1.43	228	228	0.05627	2008	Becker	Y	3.4	0.83	0.0008	3637	3637	1.18633	2012	Noyce	YY	12	0.74	0.0024	5993	18714	0.301109	499,417				
BST1 rs4038473 G/A caucasian allelic	OR	G	2009	Hasegawa	N	11.2	1.32	0.9	1.94	443	443	0.27763	2011	Do	Y	8.5	0.89	0.0017	2626	2924	0.116534	2012	Li	YY	11	0.91	0.0003	13374	4558	0.094311	1215,82				
Hormone repla therapy all	RR	O	1996	Nander	N	8.2	1.02	0.56	1.8	87	89	0.1018	2004	Ascheto	N	4.8	1.32	0.012	1249	537973	0.285179	2012	Noyce	N	9	0.9	0.67	1.21	2602	60138	0.105361	289,111			
Any family history all	OR	G	2009	Pankratz	N	4.2	0.87	0.76	1.1	857	887	0.13926	2011	Do	Y	8.5	0.92	0.012	3426	2924	0.083362	2012	Li	YY	11	0.91	0.0001	16163	52194	0.094311	1469,36				
ACADS1/MEH13 T/G rs6723108 allelic	OR	G	2009	Hasegawa	N	11.2	0.98	0.66	1.46	443	443	0.0202	2011	Do	Y	8.5	0.9	0.014	3426	2924	0.105931	2012	Li	YY	11	0.88	0.0001	13374	4723	0.122783	215,82				
TGAI rs1073761 C/T caucasian allelic	OR	G	2009	Pankratz	N	4.2	1.17	0.96	1.4	857	887	0.187	2011	PDGC USA rep	Y	8.5	1.1	0.022	2807	2215	0.09631	2012	Li	YY	23	1.52	0.0001	17121	21361	0.137672	744,191				
CDCC92HP1R rs1084784 T/G caucasian allelic	d	O	1983	Ward	Y	8.2	1.27	0.0001	49	39	2.30303	2008	Bostvedt	Y	4.6	1.85	0.0001	400	150	0.33593	2012	Rayshel	YY	30	1.52	0.0001	1675	1788	2.756974	42,847					
Obesity identification	OR	G	2009	Pankratz	Y	4.2	1.59	0.0001	857	887	0.43822	2011	Do	Y	8.5	1.2	0.0002	3426	2924	0.162392	2012	Li	YY	10	1.21	0.0001	12931	45130	0.137672	151,1					
Any family history all	RR	O	1993	Semchuk	N	8.2	5.8	0.0001	130	260	0.17978	2007	Roslan	Y	3.6	2.4	0.0003	585	228	0.879469	2012	Noyce	YY	19	4.45	0.0001	3683	544	1.492004	193,842					
BST1 rs11724938 G/A caucasian allelic	OR	G	2009	Pankratz	Y	4.2	0.78	0.0002	857	887	0.28158	2010	Hanna	Y	25.5	0.87	0.0028	2000	1985	0.139262	2012	Li	YY	28	0.88	0.0001	20034	26148	0.117813	777,077					
Obesity detection Threshold	d	O	1983	Ward	Y	8.2	0.95	0.0006	49	19	1.74125	1987	Quinn	Y	4.9	0.77	0.0001	79	40	0.306626	2012	Rayshel	YY	14	1.29	0.0001	382	621	1.338802	27,287					
LRK2 G2018S rs34637584 allelic	OR	G	2005	Nichols	Y	39	119.8	0.0008	767	1227	0.74839	2006	Kay	Y	4.6	2.21	0.0025	1425	1647	0.095978	2012	Wu	YY	39	1.46	0.0001	15879	18937	2.302715	399,462					
LRK2 rs18229 rs3394930 allelic	OR	G	2006	Ross	Y	11.2	2.03	0.001	1230	1052	0.70804	2011	Ross	N	23.9	0.62	0.36	1.07	1376	962	0.478036	2012	Wu	Y	7	1.96	0.02	4888	3789						

Breast cancer

Table with columns: Marker, # of lines, ID, Type, Year, Author, Sign, IF, OR, CI low, CI high, p, # ref, # FC, LRR1, LRR2, LRR3, LRR4, LRR5, LRR6, LRR7, LRR8, LRR9, LRR10, LRR11, LRR12, LRR13, LRR14, LRR15, LRR16, LRR17, LRR18, LRR19, LRR20, LRR21, LRR22, LRR23, LRR24, LRR25, LRR26, LRR27, LRR28, LRR29, LRR30, LRR31, LRR32, LRR33, LRR34, LRR35, LRR36, LRR37, LRR38, LRR39, LRR40, LRR41, LRR42, LRR43, LRR44, LRR45, LRR46, LRR47, LRR48, LRR49, LRR50, LRR51, LRR52, LRR53, LRR54, LRR55, LRR56, LRR57, LRR58, LRR59, LRR60, LRR61, LRR62, LRR63, LRR64, LRR65, LRR66, LRR67, LRR68, LRR69, LRR70, LRR71, LRR72, LRR73, LRR74, LRR75, LRR76, LRR77, LRR78, LRR79, LRR80, LRR81, LRR82, LRR83, LRR84, LRR85, LRR86, LRR87, LRR88, LRR89, LRR90, LRR91, LRR92, LRR93, LRR94, LRR95, LRR96, LRR97, LRR98, LRR99, LRR100. Rows include markers like NQO1 Pro337T dominant, CDKN2A rs1044396, and various other genetic variants.

Data collected for breast cancer: Data for the initial study, largest study and meta-analysis. Column c = type of study: D for "Other" and G for "Genetics". First author and year of indicated in bold correspond to initial studies that were not given by the corresponding meta-analysis.

Glaucoma

Marker	# of lines	Z1	Initial study					Largest study					Meta-analysis																							
			ES	Type	Year	Author	Sign	IF	ES	CI low	CI high	p	# pat	# HC	[L:R:5]	Year	Author	Sign	# stud	ES	CI low	CI high	p	# pat	# HC	[L:R:5]	#pat(95)									
oppreneur R425G AA+AG/GG POMA	GR	G	2002	Rezaie	N	31	6.62	0.2648	165.8076	46	100	1.8901	2003	Alward	N	3.6	0.3	0.0061	15.6449	525	162	1.204	2010	Cheng	N	10	0.89	0.65	1.21	1928	1595	0.1163	152.8			
LOXL1 SNP rs2185241 T POMA	DR	G	2007	Thorleifsson	N	31	0.83	0.63	1.09	201	198	0.1863	2008	Fan	N	2.5	0.83	0.59	1.18	323	80	0.1863	2010	Chen	N	12	0.95	0.82	1.1	2002	16372	0.05129	166.833			
LOXL1 SNP rs1048661 G POMA	DR	G	2007	Thorleifsson	N	31	0.83	0.62	1.12	199	198	0.1863	2008	Fan	N	2.5	1.02	0.7	1.51	304	80	0.1863	2010	Chen	N	10	0.93	0.84	1.03	1569	2665	0.07252	154.1			
LOXL1 SNP rs325542 G POMA	DR	G	2007	Thorleifsson	N	31	0.89	0.58	1.34	198	192	0.1278	2008	Fan	N	2.5	0.86	0.57	1.3	325	88	0.1508	2010	Chen	N	12	1.06	0.94	1.19	2210	2590	0.05827	184.167			
SPN1 gene rs38432 1 nt C and TTG	DR	G	2006	Tao	N	48	0.92	0.5155	1.632	48	49	0.0834	2007	Liu cauca	N	4.6	0.96	0.762	1.2632	280	227	0.0408	2012	Gao	N	7	1.04	0.917	1.18	1201	971	0.03922	171.571			
oppreneur M685G AA+AT/T POMA	DR	G	2002	Rezaie	Y	31	7.23			0.0001	169	422	0.9762	2003	Alward	N	3.6	1.22	0.5841	2.5014	525	162	0.1589	2010	Cheng	N	22	1.08	0.94	1.25	4996	3910	0.07696	156.182		
TP53 gene codon 72 SNP C/G allele	DR	G	2002	Lin	Y	27	2.39			0.0012	58	59	0.8713	2005	Dimas	N	5.7	1.17	0.8803	1.5636	523	178	0.187	2012	Gao P	N	9	1.1	1	1.22	1930	1463	0.09531	214.444		
TP53 gene 4892A rs180629 allele C/G	DR	G	2003	Lin	Y	18	2.72			0.0001	60	103	0.0608	2004	Fan yama	N	3.4	1	0.32	3.14	217	218	0	2013	Kim	N	10	1.44	0.958	2.18	1249	1874	0.36464	214.7		
SPN1 gene rs38432 C/T vs T/C and TTG	DR	G	2002	Aung coh1	Y	4.2	1.97			0.0369	83	100	0.6781	2007	Mabuchi	Y	3.8	2.06			0.0028	194	185	0.7227	2012	Gao	N	6	1.18	0.96	1.45	713	954	0.35551	89.125	
MTHFR gene C677T rs1801133 allele	DR	G	2002	Bleich	Y	3	3.16			0.0326	18	19	1.1566	2007	Zetterberg	N	1.1	0.82	0.61	1.1	243	187	0.1965	2013	Huo	N	10	1.17	0.94	1.46	1224	1105	0.157	222.4		
oppreneur T34T AA+AG/GG POMA	DR	G	2003	Heng	N	4.2	0.84	0.963	1.8414	185	195	0.1744	2003	Alward	N	3.6	1.09	0.769	1.6804	525	162	0.0862	2010	Cheng	Y	15	1.18			0.0137	2416	2184	0.18551	161.067		
myocilin T32G mutation POMA	DR	G	1999	Fingert	N	7.7	1.41	0.997	34.564	1693	753	0.3436	1999	Fingert	N	7.7	1.41	0.997	34.564	1693	753	0.3436	2012	Cheng	YY	12	2.17			0.002	3452	2659	0.17473	287.667		
low myopia as risk of POMA	DR	O	1999	Michell	Y	5.6	2.3			0.0045	108	3546	0.8329	2010	Kuzin	N	5.6	1.6	0.9	2.6	252	5675	0.47	2011	Marcus	YY	7	1.65			0.0003	949	28499	0.50078	135.571	
SPN1 gene rs38432 T vs C and TTG	DR	G	2002	Aung	Y	4.2	2.46			0.0098	83	100	0.6003	2007	Marcus	N	3.6	1.27	0.2853	5.7328	194	188	0.229	2012	Gao	YY	8	1.52			0.002	713	954	0.1871	81.124	
myocilin C388X mutation POMA	DR	G	1998	Alward	Y	51.7	12.73			0.0139	716	596	2.544	1999	Fingert	Y	7.7	12.84			0.0123	1693	753	2.5523	2012	Cheng	YY	9	4.68			0.0001	3820	2144	1.5483	424.444
high myopia as risk of POMA	DR	O	1999	Michell	Y	5.6	3.3			0.0004	108	3546	1.1599	2010	Kuzin	Y	5.6	2			0.0251	252	5675	0.6931	2011	Marcus	YY	7	2.46			0.0001	949	28499	0.50016	135.571
beta homocysteine and POMA	F	O	2002	Bleich	Y	3	1.43			0.0001	18	19	0.2507	2009	Michell	Y	4.6	0.65			0.0038	76	60	0.9976	2012	Huo	YY	12	0.51			0.0001	443	442	0.52564	58.9163
LOXL1 SNP rs2185241 ex60 glauc. Caucas	DR	G	2007	Thorleifsson	Y	31	2.53			0.0001	196	198	0.9282	2008	Passito	Y	3.4	2.55			0.0001	515	342	0.9381	2010	Chen	YY	12	2.35			0.0001	2156	6202	0.85442	179.667
LOXL1 SNP rs2185241 ex60 glauc. Caucas	DR	G	2007	Thorleifsson	Y	31	3.78			0.0001	198	198	1.3267	2008	Passito	Y	3.4	3.29			0.0001	512	340	1.1809	2010	Chen	YY	10	3.39			0.0001	1880	16484	1.22081	188
LOXL1 SNP rs325542 G ex60 glauc. Caucas	DR	G	2007	Thorleifsson	Y	31	27.2			0.0001	196	192	0.3032	2008	Passito	Y	3.4	3.68			0.0001	514	346	1.5507	2010	Chen	YY	13	9.3			0.0001	2224	4852	2.20001	121.071
any myopia as risk of POMA	DR	O	1994	Ponte	Y	1.5	5.56			0.0022	43	218	1.7189	1999	Wu	Y	3.4	1.48			0.0096	286	3753	0.392	2011	Marcus	YY	11	1.92			0.0001	1279	45349	0.62533	116.271

Data collected for Glaucoma: Data for the initial study, largest study and meta-analysis. Column C: type of study, O for "Other" and G for "Genetic".
 First author and year of indicated in bold correspond to initial studies that were not given by the corresponding meta-analysis

PSORIASIS

Marker	# of lines	TS	Initial study					Largest study					Meta-analysis																							
			Year	Author	Sign	IF	ES	CI low	CI high	p	# pat	# HC	[L][E][S]	Year	Author	Sign	IF	ES	CI low	CI high	p	# pat	# HC	[L][E][S]	#publist											
Apolipo E Epsilon 4 vs 2+3	OR	G	1997	Furumoto	N	2.7	0.99	0.51	1.93	100	100	0.0101	2010	Colo-Segura	N	2.7	1.28	0.89	1.83	331	400	0.2469	2013	Han	N	7	1.19	0.96	1.47	988	1092	0.133933	138			
VI D receptor polym BamI b/B allelic	OR	G	1997	Kontula	N	3.8	0.68	0.3219	1.3826	19	75	0.4155	2012	Rucevic	N	1.7	1.27	0.9842	1.6843	180	366	0.239	2013	Liu	N	8	1.07	0.901	1.277	644	927	0.067659	80.5			
PTN22 C165S1 rs2076601 T/C allelic	OR	G	2005	Criswell	N	11	0.91	0.44	1.89	63	2064	0.0943	2008	Smith	N	3.8	1.1	0.79	1.54	682	2650	0.0553	2012	Chen	N	10	1.15	1	1.33	3758	7987	0.139162	375.6			
VI D receptor polym Apal a/A allelic	OR	G	1998	Park	Y	6.2	0.43			104	104	0.844	2012	Rucevic	N	1.7	1.09	0.8501	1.4118	180	366	0.0862	2013	Liu	N	8	1.03	0.888	1.191	715	937	0.029559	89.375			
VI D receptor polym TagI t/vT allelic	OR	G	2002	Okita	N	2.7	1.4	0.6074	3.2031	50	86	0.3365	2006	Halsall	N	3.6	0.7	0.5008	1.0533	205	80	0.3567	2013	Liu	Y	7	1.24			0.013	704	805	0.151511	100.5714		
IL23R rs2201841 G/A	OR	G	2007	Capon	N	4.2	1.1	0.681	1.424	312	287	0.0553	2010	Sun	N	3.5	0.94	0.882	1.0076	3163	6722	0.0619	2012	Zhu c	Y	13	1.12			0.068	10278	13884	0.13324	190.6515		
VEGF +405 G/C allelic	OR	G	2004	Young (severe)	Y	6.2	0.59			72	102	0.5276	2007	Butt	N	1.9	1.08	0.77	1.47	258	154	0.0583	2010	Wu	Y	7	1.17			0.006	1268	1513	0.157004	180.8571		
Apolipo E Epsilon 2 vs 3+4	OR	G	1997	Furumoto	Y	2.7	1.94			0.0184	100	100	0.6627	2010	Colo-Segura	N	2.7	1.28	0.94	1.75	331	400	0.2469	2013	Han	YY	7	1.47			0.0001	968	1092	0.385262	138	
Alcohol consumption	OR	O	1985	Chaput	Y	17	0.14			0.0001	44	1943	1.9601	2007	Muertes	N	4.8	0.93	0.85	1.03	2582	6332	0.0728	2012	Zhu b	YY	15	1.53			0.002	7891	17154	0.025268	612.0667	
IL23R rs7530511 C/T	OR	G	2007	Cargill-all	Y	11	0.78			0.014	1448	1382	0.2485	2008	Nair	N	6.2	0.85	0.7271	1.0032	1444	1395	0.1625	2012	Zhu c	YY	12	0.82			0.0001	6882	9086	0.198451	171.8333	
Smoking asian	OR	O	2002	Zhang	Y	1.3	2.33			0.0001	440	453	0.8459	2010	Zohren	Y	3.5	1.54			0.0001	6516	12197	0.4318	2011	Zhu	YY	9	2.01			0.0001	18253	28246	0.698133	2028.111
Apolipo E Epsilon 3 vs 2+4	OR	G	1997	Furumoto	Y	2.7	0.66			0.0497	100	100	0.4463	2010	Colo-Segura	Y	2.7	0.75			0.0005	331	400	0.2744	2013	Han	YY	7	0.71			0.0001	968	1092	0.342459	138
HLA-Cw*06	OR	G	1978	Bremer	Y	2.7	8.28			0.0001	77	125	2.1198	2003	Sudrensson	Y	3.8	6.07			0.0001	1006	512	1.8034	2011	Wu	YY	18	4.56			0.0001	3419	3597	1.151517	189.8444
IL23R rs11295026 A/G	OR	G	2007	Cargill-all	Y	11	0.63			0.0002	1448	1382	0.462	2008	Nair	Y	6.2	0.78			0.0245	1442	1417	0.2465	2012	Zhu c	YY	17	0.61			0.0001	8869	18143	0.494298	521.7059
Smoking european	OR	O	1992	Wills	Y	3.8	2.78			0.0007	108	100	1.0223	2009	Prodanowich	Y	4.8	3.4			0.0001	3236	2500	1.2226	2011	Zhu	YY	16	1.89			0.0001	7280	14012	0.688133	453.75

Data collected for Psoriasis: Data for the initial study, largest study and meta-analysis. Column C : type of study, O is for "Other" and G for "Genetic".
 First author and year of indicated in bold correspond to initial studies that were not given by the corresponding meta-analysis

Rheumatoid Arthritis

Marker	# of times	ES	Year	Initial study	N	P	ES	CI low	CI high	p	# cases	# HC	OR	95% CI	P-value	# cases	# HC	OR	95% CI	P-value	# cases	# HC	OR	95% CI	P-value	# cases	# HC	OR	95% CI	P-value	# cases	# HC	OR	95% CI	P-value					
																																				ES	CI low	CI high	p	OR
CTLA-4 +49 rs2171775 G/A allelic G vs A	CR	G	1998	Seldin	N	27	1.15	0.9236	1.4358		298	456	0.1392	0.2689		2008	Ishihara	N	9.1	1.14	0.9774	1.3313		1490	448	0.13103	0.2176		2012	Li	Y	19	1.11		0.01	7540	8231	0.10343	0.16146	0.74283
CD134 rs1034569 A/G allelic A vs G	CR	G	1999	Hlatky	N	91	0.87	0.659	1.168		276	277	0.1391	0.2689		2012	Bisanz	N	8	1.14	0.914	1.419		2291	2731	0.11320	0.16146	0.74283												
IL-18 rs6747C rs1584518 allelic C vs G	CR	G	2003	Swatigam	N	27	0.85	0.9901	1.244		74	225	0.1503	0.2176		2012	Chen b	N	7	0.75	0.562	1.006		1958	1707	0.28762	0.16146	0.74283												
CTLA-4 rs1704 rs3087243 A/G allelic A vs G	CR	G	2004	Barton	N	75	0.96	0.8345	1.1169		719	795	0.0428	0.0001		2005	Plenge	N	11.2	0.95	0.8428	1.0731		1505	878	0.05129	0.16146	0.74283												
IL-10 rs1028 rs180292 G/A allelic G vs A	CR	G	1999	Coakley	N	42	0.93	0.6457	1.3434		117	295	0.0724	0.0001		2010	Pradnyasaka	Y	1.3	0.98				244	106	2.4451	0.16146	0.74283												
IL-10 rs1028 rs180292 G/A allelic G vs A	CR	G	1999	Coakley	N	42	0.95	0.702	1.2884		117	295	0.0513	0.0001		2004	Pradyakov b	N	3.3	0.76	0.5553	1.0085		284	286	0.28768	0.16146	0.74283												
IL-1RN rs11784235 G/A allelic G vs A	CR	G	1999	Berman	N	37	0.93	0.5223	1.6938		43	100	0.0726	0.0002		2006	Carreira	Y	7.5	1.46				441	287	0.37844	0.16146	0.74283												
OST1 rs1049904 G/A allelic G vs A	CR	G	1999	Layton	N	42	1.32	0.79	2.3		81	595	0.2773	0.0001		2006	Pradyakov	N	7.5	1	0.8	1.24		1733	1928	0	0.16146	0.74283												
TNFR3 rs122897 G/A allelic G vs A	CR	G	2008	Rusada	N	75	1.04	0.88	1.232		724	542	0.0352	0.0001		2008	Kim	N	3.3	0.97	0.863	1.102		1034	979	0.03346	0.16146	0.74283												
IL-18 rs11111054 C/G vs G/A vs A	CR	G	1999	Chambers	N	75	1.08	0.98	2.88		168	124	0.1818	0.0001		2006	Lotrion	N	7.5	1.09	0.93	1.26		1277	1101	0.06168	0.16146	0.74283												
RFX5 rs12932 G/A allelic G vs A	CR	G	2007	Sjogstrand-1	N	75	1.05	0.922	1.189		1530	861	0.0485	0.0001		2007	Sjogstrand-1	N	7.5	1.05	0.922	1.189		1530	861	0.04875	0.16146	0.74283												
Smoking past	CR	G	1987	Wessely	N	3.1	1.15	0.54	2.43		78	522	0.1398	0.0001		1999	Karlson	N	7.5	1.06	0.96	1.13		7697	369784	0.06827	0.16146	0.74283												
CD14 rs104322221 G/A allelic G vs A	CR	G	1999	Coakley	N	75	0.68	0.448	1.034		278	285	0.3857	0.0001		2007	Lunnen	N	2.5	0.8	0.71	1.191		836	645	0.10588	0.16146	0.74283												
Smoking ever/never	CR	G	1987	Wessely	N	3.1	1.5	0.93	2.44		78	522	0.4055	0.0001		1999	Karlson	Y	7.5	1.13				7697	369784	0.12222	0.16146	0.74283												
FCGR3B copy # c2	CR	G	2009	Breunis	N	5.2	1.31	0.51	3.34		112	129	0.27	0.0003		2010	McKinney (Dutch)	Y	9.1	2.01				741	701	0.69813	0.16146	0.74283												
RFX5 rs1034569 G/A allelic G vs A	CR	G	2006	Rusada	N	75	0.9	0.766	1.052		724	542	0.1051	0.002		2007	Sjogstrand	Y	7.5	0.85				1030	861	0.16261	0.16146	0.74283												
TRAF1 rs12187847 G/A allelic G vs A	CR	G	2007	Plenge b (NARAC1)	Y	51.7	1.43			0.0001	908	1260	0.3577	0.0001		2008	Barton (UK)	N	7.7	1.03	0.97	1.09		5063	3849	0.02956	0.16146	0.74283												
PADI4 rs1049904 G/A allelic G vs A	CR	G	2003	Suzuki	Y	35.2	1.4			0.0001	853	650	0.3365	0.0001		2010	Burn T	N	9.1	1.02	0.96	1.09		3732	3039	0.01198	0.16146	0.74283												
TRAF1 rs12187847 G/A allelic G vs A	CR	G	2007	Plenge b (NARAC1)	Y	51.7	1.5			0.0001	908	1260	0.4055	0.0001		2007	Postler (ARC)	N	9.1	1	0.78	1.27		1469	738	0	0.16146	0.74283												
PTPN22 rs1881 rs2479601 T/C allelic T vs C	CR	G	2004	Bergoch-2	Y	11.2	2.13			0.0001	840	928	0.7561	0.0001		2008	Seldin	Y	7.3	1.47				1030	1490	0.38528	0.16146	0.74283												
TNFR3 rs122897 G/A allelic G vs A	CR	G	2007	Thompson	Y	35.2	1.22			0.0001	4806	3695	0.201	0.0001		2007	Thompson	Y	35.2	1.23				4806	3695	0.20701	0.16146	0.74283												
FCRL3 rs169792884 C/T allelic C vs T	CR	G	2008	Kochi	Y	35.2	1.37			0.0001	834	649	0.3148	0.0001		2006	Newman	Y	7.5	1.19				1187	462	0.17355	0.16146	0.74283												
PADI4 rs1049904 G/A allelic G vs A	CR	G	2003	Suzuki	Y	35.2	1.31			0.0005	823	650	0.27	0.0003		2010	Ikeai	Y	7.5	1.21				1230	848	0.19902	0.16146	0.74283												
PADI4 rs1049904 G/A allelic G vs A	CR	G	2003	Suzuki	Y	35.2	1.3			0.0006	823	650	0.2624	0.0004		2008	Barton b	N	7.5	1.09	0.92	1.28		839	481	0.08618	0.16146	0.74283												
TNFR3 rs122897 G/A allelic G vs A	CR	G	2007	Plenge	Y	35.2	0.69			0.002	873	1413	0.3711	0.0014		2010	Shimane	Y	7.5	1.29				10014	3415	2.926	0.25464	0.16146												
IL-23 rs1943151 A/G allelic A vs G	CR	G	2007	WTCDC	Y	38.6	1.13			0.0058	1859	2926	0.1222	0.0058		2012	WTCDC	Y	38.6	1.13				1859	2926	0.12222	0.16146	0.74283												
FCGR3A rs1897 F/F vs V/V	CR	G	2000	Nieto	Y	7.3	1.98			0.0079	117	142	0.0631	0.0079		2010	Robinson	N	9.1	0.93	0.824	1.045		2993	1731	0.07291	0.16146	0.74283												
IL-23 rs1049904 G/A allelic G vs A	CR	G	2007	WTCDC	Y	38.6	1.12			0.0084	1859	2926	0.1133	0.0084		2012	Seng b	Y	7	1.08				1859	2926	0.11333	0.16146	0.74283												
PADI4 rs1049904 G/A allelic G vs A	CR	G	2003	Suzuki	Y	35.2	1.5			0.0119	288	181	0.4055	0.0119		2011	Kochi b	Y	9.1	1.14				1019	807	0.3103	0.16146	0.74283												
MIR27A rs63433305 G/A allelic A vs G	CR	G	2005	Swanberg	Y	35.2	0.85			0.0121	1262	1599	0.1623	0.0121		2006	Eye	N	7.5	0.96	0.868	1.073		1401	2476	0.04082	0.16146	0.74283												
STAT4 rs7574895 T/G allelic T vs G	CR	G	2007	Remmers (ERA)	Y	51.7	1.18			0.0186	1529	881	0.1655	0.0186		2006	Barton (UK)	N	7.7	1.17				3299	3024	0.157	0.16146	0.74283												
TGFB1 rs1892873 G/A allelic G vs A	CR	G	2002	Sjogren	Y	5.1	1.86			0.0341	155	110	0.0201	0.0341		2009	Phouasat	N	4.2	0.8	0.61	1.35		395	401	0.00538	0.16146	0.74283												
SLE22A4 rs2421 F2 allelic F vs 2	CR	G	2003	Takahara	Y	35.2	1.28			0.049	819	656	0.2231	0.049		2009	Plenge swedish	N	11.2	1.13	0.9	1.42		1916	874	0.12222	0.16146	0.74283												
Smoking current	CR	G	1987	Wessely	Y	3.1	1.66			0.049	78	522	0.2006	0.049		1999	Karlson	Y	7.5	1.18				7697	369784	0.16591	0.16146	0.74283												

Data collected for Rheumatoid Arthritis: Data for the initial study, largest study and meta-analysis. Column C: type of study, O is for "Other" and G for "Genetic".
 First author and year of indicated in bold correspond to initial studies that were not given by the corresponding meta-analysis

Annexe 2

RESEARCH ARTICLE

Poor replication validity of biomedical association studies reported by newspapers

Estelle Dumas-Mallet^{1,2*}, Andy Smith¹, Thomas Boraud², François Gonon²

1 Centre Emile Durkheim, CNRS UMR5116 at Université de Bordeaux, Pessac, France, **2** Institute of Neurodegenerative Diseases, CNRS UMR5293 at Université de Bordeaux, Bordeaux, France

* estelle.mallet@u-bordeaux.fr

Abstract

Objective

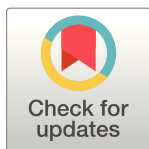
To investigate the replication validity of biomedical association studies covered by newspapers.

Methods

We used a database of 4723 primary studies included in 306 meta-analysis articles. These studies associated a risk factor with a disease in three biomedical domains, psychiatry, neurology and four somatic diseases. They were classified into a lifestyle category (e.g. smoking) and a non-lifestyle category (e.g. genetic risk). Using the database *Dow Jones Factiva*, we investigated the newspaper coverage of each study. Their replication validity was assessed using a comparison with their corresponding meta-analyses.

Results

Among the 5029 articles of our database, 156 primary studies (of which 63 were lifestyle studies) and 5 meta-analysis articles were reported in 1561 newspaper articles. The percentage of covered studies and the number of newspaper articles per study strongly increased with the impact factor of the journal that published each scientific study. Newspapers almost equally covered initial (5/39 12.8%) and subsequent (58/600 9.7%) lifestyle studies. In contrast, initial non-lifestyle studies were covered more often (48/366 13.1%) than subsequent ones (45/3718 1.2%). Newspapers never covered initial studies reporting null findings and rarely reported subsequent null observations. Only 48.7% of the 156 studies reported by newspapers were confirmed by the corresponding meta-analyses. Initial non-lifestyle studies were less often confirmed (16/48) than subsequent ones (29/45) and than lifestyle studies (31/63). Psychiatric studies covered by newspapers were less often confirmed (10/38) than the neurological (26/41) or somatic (40/77) ones. This is correlated to an even larger coverage of initial studies in psychiatry. Whereas 234 newspaper articles covered the 35 initial studies that were later disconfirmed, only four press articles covered a subsequent null finding and mentioned the refutation of an initial claim.



OPEN ACCESS

Citation: Dumas-Mallet E, Smith A, Boraud T, Gonon F (2017) Poor replication validity of biomedical association studies reported by newspapers. PLoS ONE 12(2): e0172650. doi:10.1371/journal.pone.0172650

Editor: Jelte M. Wicherts, Tilburg University, NETHERLANDS

Received: October 27, 2016

Accepted: February 7, 2017

Published: February 21, 2017

Copyright: © 2017 Dumas-Mallet et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Conclusion

Journalists preferentially cover initial findings although they are often contradicted by meta-analyses and rarely inform the public when they are disconfirmed.

Introduction

Over recent years many scientific editorials have highlighted the "reproducibility crisis" of biomedical findings [1–5]. Empirical studies have actually observed that many initial findings are refuted by subsequent studies, a trend that holds true at all three levels of biomedical research: i) preclinical studies [6, 7], ii) associations between biomarkers or risk factors and diseases [8–11], and iii) clinical trials [12–16]. These empirical studies, as well as theoretical considerations [1], have highlighted that initial findings on a novel scientific question are especially vulnerable to refutation. However, others have argued that it is a mistake to expect 100% replication [17]: "a single study almost never provides definitive resolution for or against an effect and its explanation." Scientific progress is a cumulative process that reduces uncertainty by combining uncertain initial findings with validation by subsequent studies. In other words, the low replication validity of many initial studies does not hinder the advancement of knowledge as long as science remains self-corrective [18].

However, an unrecognized consequence of the low replication validity of single studies concerns their media coverage. In the case of attention deficit hyperactivity disorder (ADHD), we previously found that newspapers preferentially report initial findings although most are refuted or strongly attenuated by subsequent studies [19]. We examined the replication validity of ten biomedical studies related to ADHD, which were widely reported by the media during the 1990's. Among them, seven were found to be initial studies and all seven were either refuted or strongly attenuated by subsequent studies [19]. However, the strength of evidence provided by this preliminary study was limited by its sample size. The present study examined the preferential media coverage of early biomedical findings on a larger scale. Another weakness of our preliminary study was related to the heterogeneity of the ten ADHD studies that were examined: one was a pre-clinical study in mice, another estimated ADHD prevalence, three investigated the effectiveness of medication and five investigated the association of ADHD with biomarkers or risk factors. In the present study we did not consider pre-clinical studies because they are rarely included in meta-analyses, thus precluding assessment of their replication validity. We also excluded clinical trials investigating treatment effectiveness because their publication is influenced by stakeholders, in most cases pharmaceutical companies [20, 21], and because their media coverage is biased by this influence [22, 23]. Because this influence represents a parameter that might confound our analysis, we decided to focus on studies associating biomarkers and risk factors with diseases. This type of biomedical research is crucial for improving the understanding, diagnosis and prevention of diseases and is thus often reported by newspapers.

We recently built a large database of 595 meta-analyses investigating biomarkers and risk factors associated with four psychiatric disorders, four neurological pathologies and four somatic diseases [11]. Starting from the assumption that each meta-analysis provided the best estimate of the 'true' effect size, we showed that most effect sizes reported by initial studies were inflated. In the present study we investigated three questions. First, we quantified and characterized the newspaper coverage of the scientific publications included in our database. Second, we tested whether scientific observations reported by newspapers were confirmed,

attenuated or refuted by the corresponding meta-analyses. We explored several factors that might influence the percentage of studies covered by newspapers and the replication validity of scientific observations that are covered: the type of study (initial versus subsequent studies and null findings versus "positive" ones), the prestige of the scientific journal that published them, the biomedical domain and the type of association. In particular, as previous studies showed that newspapers preferentially cover studies associating diseases with lifestyle risk factors (e.g. smoking) [24, 25], we considered this type of association to be a category. Third, we analyzed the possible media coverage of subsequent studies contradicting initial claims that had previously been reported by newspapers.

Methods

Scientific studies and their classification

The database described in our previous publication [11] included 663 meta-analyses investigating the association between biomarkers, risk factors or cognitive/behavioral observations and twelve pathologies belonging to three biomedical domains: psychiatry (Attention Deficit Hyperactivity Disorder (ADHD), autism, major depression, and schizophrenia), neurology (Alzheimer's and Parkinson's diseases, epilepsy, multiple sclerosis) and four somatic diseases (breast cancer, glaucoma, psoriasis and rheumatoid arthritis). In a preliminary step of the present study we observed that initial studies related to cognitive/behavioral observations were never covered by newspapers. Therefore, the 68 meta-analyses dealing with cognitive/behavioral observations were not considered in the present study. The 595 meta-analyses considered here were reported by 306 meta-analysis articles and included primary observations reported in 4775 scientific articles. As described previously, initial studies were distinguished from subsequent ones on the basis of their publication date (their references are given in our previous study [11]). From this database we excluded primary articles published before 1981. Indeed, we used the *Dow Jones Factiva* database to test whether these primary articles were covered by newspapers and the *Factiva* database starts on January 2nd 1981. Thus our final database includes 5029 scientific articles: 405 initial studies, 4218 subsequent studies and 306 meta-analysis articles.

In our previous study, these 595 meta-analyses were classified into three types: genetic studies, brain imaging studies and "other" studies. This latter type included a wide variety of biochemical, biophysical and epidemiological studies. Previous studies [24, 25] and a preliminary step of the present study led us to subdivide all association studies into two categories: a lifestyle category corresponding to any association between a pathology and a risk factor on which each subject can individually decide to act (e.g. red meat consumption, smoking, non compulsory vaccination) and a non-lifestyle category.

Impact factor of the journals that published scientific studies

We used the *Journal Citation Reports* from *The Web of Science* to determine the 2012 impact factor of the journals that published the 5029 articles of our database. In our previous study we observed that studies associating ADHD with a biomarker or a risk factor were rarely covered by newspapers if they were published in a scientific journal with an impact factor below five [19]. Therefore, when testing whether each study of our database was echoed by newspapers, we limited our initial search to articles published in scientific journals with an impact factor equal or superior to five. We observed that among the 91 scientific articles reported by newspapers and related to other pathologies than breast cancer, the lowest impact factor was 6.5. We found no press coverage for the 220 articles published in journals with an impact factor between 5 and 6.5. Therefore, it seems unlikely that the 2342 primary studies not related to

breast cancer and published in journals with an impact factor below five attracted journalists' attention. As a result, we did not check the media coverage of these 2342 primary studies. Regarding breast cancer, among the 64 scientific articles reported by newspapers, we found 14 articles with an impact factor between 5 and 6.5. Therefore, we extended our search for a possible media coverage of breast cancer studies with an impact factor between 4 and 5. However, we found no media coverage for these 340 articles. Therefore, we did not check the media coverage of the 339 primary studies related to breast cancer with an impact factor below four.

In order to test the influence of the impact factor on media coverage, we considered three classes. Prestigious journals with an impact factor ≥ 30 (*Nature*, *Science*, *JAMA*, *The Lancet* and *The New England Journal of Medicine*) formed the first class. The limit between the two other classes was adjusted in order to keep the number of covered articles roughly equal in both classes. Thus, the second class included articles published in journals with an impact factor between 10 and 30 and the third one all other articles ($IF \leq 10$).

Newspaper coverage of scientific studies

As previously described [19], we used the *Dow Jones Factiva* database to find newspaper articles echoing the scientific studies of our database. Briefly, the following keywords were used in each search: "study OR research* OR scientist*". Moreover, each search also used specific keywords that included the name of the pathology, one or several keywords that characterized the study (e.g. "gene", "milk", "birth weight") and was restricted to one month after the publication date of the study. Some searches, particularly about genetic associations, led to many items. In those cases, we added the name of the authors of the study, or of the university to the keywords. We only considered newspaper articles written in English and published in the general press. Articles published in the specialized press (e.g. *Pharma Business Week*), as well as articles published by any press agency (e.g. *Associated Press*), were not taken into account. Sometimes newspaper articles found by this search actually highlighted a biomedical finding that was not included in our database of 595 meta-analyses. For example, some newspaper articles underlined risk factors associated with colon cancer, but did not mention the association of the same risk factor with breast cancer although it was also reported in the corresponding scientific article. In these few cases this media coverage was not considered. This search was performed by one author (EDM). Another author (FG) independently checked the possible media coverage of 20% of scientific studies. Both authors always reached agreement on whether any scientific study was echoed or not. A few minor disagreements occurred regarding the number of newspaper articles reporting some studies and they were resolved by discussion.

Replication validity of scientific observations reported by newspapers

We used our previously published database of 595 meta-analyses to check whether biomedical findings echoed by newspapers were confirmed, or not, by the corresponding meta-analyses. As previously described [11], a primary observation was considered in agreement with the corresponding meta-analysis if it satisfied two criteria: first, both observations were in nominal agreement (i.e. both reported a statistically significant effect in the same direction or the absence of a significant effect) and, second, when both reported a significant effect, the effect size of the primary observation must not be more than twice as large as that of the corresponding meta-analysis (i.e. effect size inflation below 100%). Sometimes single scientific articles reported several associations included in our 595 meta-analyses while newspaper articles that echoed them did not specify which association they highlighted. This is often the case with genetic studies: newspapers reported that new genes have been found to be associated with a disease, whereas the scientific article described significant associations with several single

nucleotide polymorphisms. In these cases we considered the replication validity of the most credible association (i.e. with the smallest *p* value of the statistical test). Thus, we assessed the replication validity of any scientific article covered by newspapers under the most favorable conditions.

Results

Overview of scientific articles reported by newspapers

Among the 4723 primary studies of our database 156 were echoed by at least one newspaper article (Table 1). These 156 studies were covered by 1475 newspaper articles and were related to ten pathologies (ADHD, autism, major depression, schizophrenia, Alzheimer’s and Parkinson’s diseases, multiple sclerosis, breast cancer, glaucoma and rheumatoid arthritis). The primary studies included in our database and related to epilepsy or psoriasis did not attract newspaper attention. Moreover, 86 newspaper articles covered five out of the 306 meta-analysis articles of our database. The references of these 156 studies and of the corresponding meta-analysis articles are given in Supporting Information (S1 Text). The raw data about the studies covered by newspapers and their corresponding meta-analyses are also given (S1 File).

The number of newspaper articles echoing each of these 161 scientific articles (156 primary and 5 meta-analysis articles) ranged from 1 to 50 and this number was positively correlated with the impact factor of the journal that published them (linear regression, $R^2 = 0.107$, F-test on variance equality: $p = 0.003$, Pearson’s test: $p < 0.0001$).

Biomedical domain and association type of covered studies

The percentage of primary articles covered by newspapers ranged from 1.74% to 10.5% depending on the biomedical domain and the association type (Table 1). Whatever the domain, lifestyle studies were more often covered than non-lifestyle ones (Table 1). Compared to this major effect of association type, the coverage rate does not strongly differ between biomedical domains (Table 1). Newspapers also covered a larger proportion of meta-analysis articles of the lifestyle category (Table 1). Altogether, newspapers echoed 161 scientific articles, of

Table 1. Number of primary articles included in meta-analyses and number of primary articles covered by newspapers classified by domains and association types.

Primary articles	All	lifestyle	non-lifestyle
PSY	1905	63	1842
NEURO	1279	158	1121
SOMA	1539	418	1121
All	4723	639	4084
<i>Covered articles</i>			
PSY	38 (1.99)	6 (9.52)	32 (1.74)
NEURO	41 (3.21)	13 (8.23)	28 (2.50)
SOMA	77 (5.0)	44 (10.53)	33 (2.94)
All	156 (3.30)	63 (9.86)	93 (2.28)
Meta-analysis articles			
all	306	37	269
covered	5 (1.63)	2 (5.41)	3 (1.11)

PSY: psychiatry, NEURO: Neurology, SOMA: somatic diseases, lifestyle: association between a pathology and a risk factor on which each subject can act, non-lifestyle: any other association studies. The percentage of articles covered by newspapers is indicated in parentheses.

doi:10.1371/journal.pone.0172650.t001

which 65 (40.4%) were of the lifestyle category. Moreover, lifestyle associations tended to receive a larger newspaper coverage (mean: 11.9 newspaper articles per study) than non-lifestyle associations (mean: 8.2 newspaper articles per study). However, this difference does not reach statistical significance. Overall, this larger press coverage of lifestyle studies was not related to the impact factor of the scientific journals that published them. Indeed, lifestyle scientific studies were published in less prestigious journals (median IF = 10.6) than those of the non-lifestyle category (median IF = 17.6) and this difference is statistically significant (unpaired t test, $p = 0.0003$).

Media coverage of initial studies versus subsequent ones

Initial studies attract more media attention than subsequent studies. Indeed, 13.1% of initial studies were covered by newspapers whereas only 2.4% of subsequent studies were (raw data are given in Supporting Information: [S2 Text](#)). This preferential coverage is largely influenced by the type of association and the impact factor of the scientific journal. Among the 639 primary articles of the lifestyle category, 39 reported initial findings and 600 subsequent observations. Newspapers almost equally covered these initial findings (5 covered articles, 12.8%) and these subsequent observations (58 covered articles, 9.7%). In contrast, among the 4084 primary articles investigating non-lifestyle associations, newspapers preferentially covered initial findings ([Fig 1](#)). Indeed, they echoed 48 of the 366 initial non-lifestyle studies (13.1%), but only 45 of the 3718 subsequent ones (1.2%). This difference reaches statistical significance (Chi2 test: $X^2 = 211.52$ $p < 0.0001$). This preferential coverage of initial non-lifestyle studies also holds true when considering articles published in scientific journals with an impact factor below 10 (Chi 2 test: $X^2 = 17.73$ $p < 0.0001$) and with an impact factor between 10 and 30 (Chi 2 test: $X^2 = 29.11$ $p < 0.0001$) ([Fig 1](#)). However, for the 83 non-lifestyle articles published in prestigious journals, no significant preferential coverage of initial observations was observed (Chi 2 test: $X^2 = 3.49$ $p = 0.062$). Finally, the percentage of articles covered by newspapers strongly increased with the impact factor ([Fig 1](#)). This effect is statistically significant regarding initial or subsequent non-lifestyle studies (2x3 Chi 2 tests: $X^2 = 124.4$ $p < 0.0001$ or $X^2 = 612.1$ $p < 0.0001$, respectively) and for lifestyle studies (2x3 Fisher test: $p < 0.0001$). Therefore, publishing a study in a prestigious journal considerably increases its chance of being covered by newspapers irrespective of the type of association investigated or its newness.

Media coverage of "positive" versus null findings

All 53 initial studies covered by newspapers reported a "positive" finding (i.e. a statistically significant effect). It follows that none of the 174 initial studies included in our database and reporting a null effect (i.e. 43% of all initial studies) attracted journalists' attention ([Table 2](#)). Among the 103 subsequent studies covered by newspapers, only 14 studies reported a null effect. Regarding the non-lifestyle associations, among the 45 subsequent studies covered by newspapers, only five reported a null finding. Moreover this media coverage was small (10 articles) compared to that of all non-lifestyle studies (742 newspaper articles). The 58 subsequent studies of the lifestyle category were echoed by 733 newspaper articles. Only nine subsequent studies reported the absence of a significant association and these null findings were reported by 65 newspaper articles. Therefore, assuming that, in our database, the percentage of subsequent studies reporting a null finding is at least as large as that of initial studies (43%) we observed that newspapers under-reported them. Indeed, only 13.6% of the subsequent studies covered by newspapers reported a null finding ([Table 2](#)). Moreover, the number of newspaper articles covering each null finding was small, particularly regarding non-lifestyle studies ([Table 2](#)). Finally, only one of the five meta-analysis articles covered by newspapers reported a

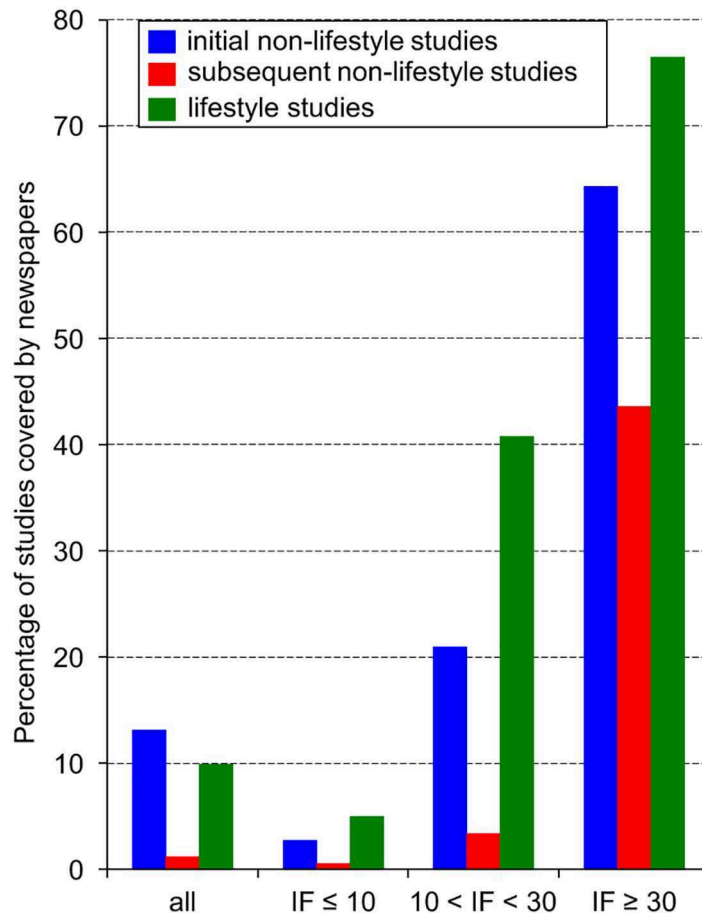


Fig 1. Preferential coverage of initial findings and influence of the impact factor (IF). The figure shows the percentage of primary studies that are covered by newspapers depending on the study type (lifestyle versus non-lifestyle). Studies of the lifestyle category described associations linking a pathology to a risk factor on which each subject can act. Regarding non-lifestyle articles, the figure also contrasts initial articles with subsequent ones. Differences in the media coverage between initial studies and subsequent ones were statistically significant (see text) except for studies published in prestigious journals ($IF \geq 30$). Raw data are given in Supporting Information (S2 Text).

doi:10.1371/journal.pone.0172650.g001

null finding. Moreover, this null finding was echoed by only four newspaper articles whereas the four "positive" meta-analyses attracted 82 press articles.

Replication validity of primary studies reported by newspapers

Overall, among the 156 primary studies echoed by newspapers 76 (48.7%) reported a main finding consistent with the corresponding meta-analysis (raw data are given in Supporting Information: S2 Text). Among the 156 primary studies covered by newspapers 53 were initial studies of which only 18 (34.0%) were confirmed by the corresponding meta-analyses. In contrast, of the 103 subsequent studies 58 (56.3%) were confirmed and this difference is statistically significant (Chi2 test: $X^2 = 6.99$ $p = 0.0082$). Regarding the 63 lifestyle studies echoed by newspapers, 49.2% of their findings were confirmed by the corresponding meta-analyses without any effect of the impact factor (Fig 2). Only 33.3% of the main findings reported by the 48 initial non-lifestyle studies and covered by newspapers were actually validated by the corresponding meta-analyses. This replication rate was not influenced by the impact factor (Fig 2).

Table 2. Newspaper coverage of primary studies reporting null findings.

	all	lifestyle	non-lifestyle
all initial studies in the database	405	41	364
"positive" findings	231	21	210
null findings	174	20	154
(% of null findings)	(43%)	(48.8%)	(42.3%)
initial studies covered:			
all	53	5	48
null findings	0	0	0
subsequent studies covered:			
all	103	58	45
null findings	14	9	5
Newspaper articles:			
all	1475	733	742
covering null findings	75	65	10

doi:10.1371/journal.pone.0172650.t002

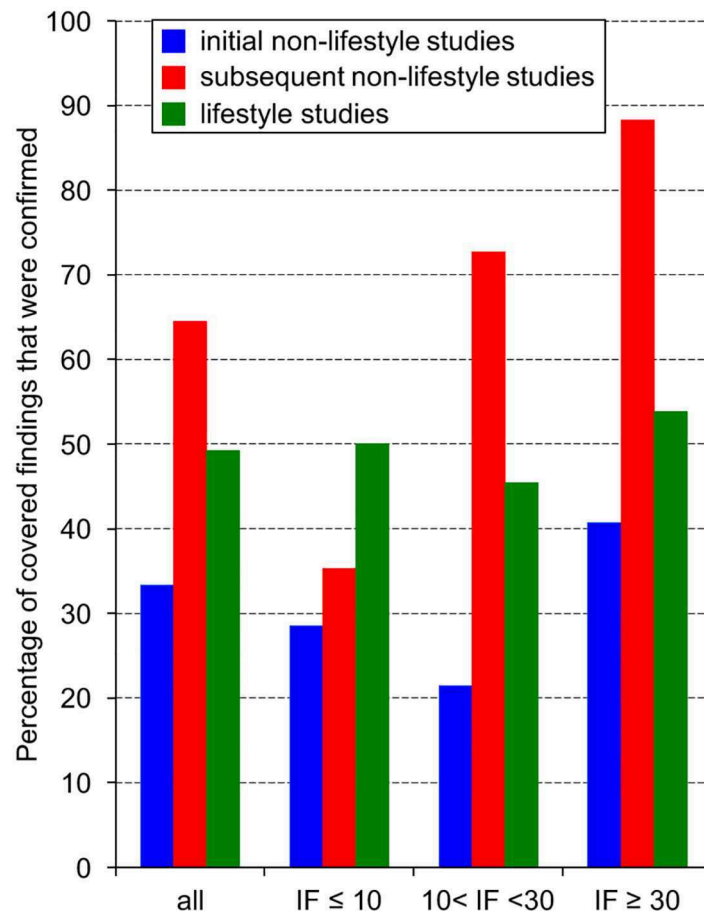


Fig 2. Replication validity of primary articles reported by newspapers. The figure shows the percentage of primary articles echoed by newspapers whose main finding was consistent with the corresponding meta-analysis. We considered here the same three categories as in Fig 1: primary articles of the lifestyle category and initial or subsequent non-lifestyle studies. Raw data are given in Supporting Information (S2 Text).

doi:10.1371/journal.pone.0172650.g002

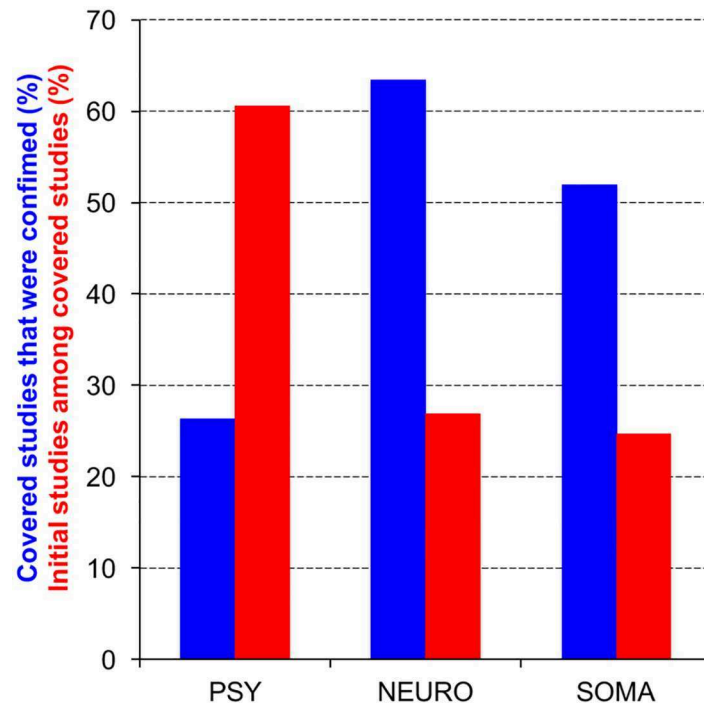


Fig 3. Replication validity of primary studies reported by newspapers in three biomedical domains. The blue bars show the percentage of primary studies covered by newspapers whose main finding was consistent with the corresponding meta-analysis. The red bars show the percentage of initial findings among primary studies echoed by newspapers and related to four psychiatric disorders (PSY), four neurological diseases (NEURO) and four somatic diseases (SOMA). Raw data are given in Supporting Information (S2 Text).

doi:10.1371/journal.pone.0172650.g003

In contrast the 45 subsequent non-lifestyle studies covered by newspapers were more often consistent with the corresponding meta-analyses than the initial non-lifestyle studies (Fig 2). Their replication rate was indeed larger than that of initial studies whether considering them as a whole (64.4% versus 33.3% for initial studies; Chi2 test, $X^2 = 9.0$ $p = 0.0027$) or when considering only the 17 studies published in prestigious journals (88.2% versus 40.7% for initial studies; Fisher test, $p = 0.0021$). The replication rate of subsequent non-lifestyle studies increased with the impact factor (2x3 Fisher test, $p = 0.0055$).

When we classified the 156 primary studies echoed by newspapers according to the biomedical domain (Fig 3), studies related to psychiatric disorders were less often confirmed by the corresponding meta-analyses than those related to neurological diseases (chi2 test; $X^2 = 10.94$ $p = 0.0009$) and to somatic diseases ($X^2 = 6.8$ $p = 0.0091$). Moreover, studies related to psychiatry and covered by newspapers were more often initial findings than those in neurology ($X^2 = 9.13$ $p = 0.0025$) and for somatic diseases ($X^2 = 14.11$ $p = 0.0002$). The low replication validity of studies in psychiatry appears thus to be correlated with the fact that newspapers favor initial findings in this domain even more than in other domains. This correlation further supports our previous conclusion [19]: the replication validity of biomedical studies covered by newspapers is poor because journalists favor initial findings over subsequent studies.

Analysis of primary studies covered by three newspapers or more

When considering only the 102 primary studies covered by at least three newspaper articles, we observed the same patterns (raw data are given in Supporting Information: S2 Text).

Journalists covered initial studies more often (32/405) than subsequent studies (70/4318). In particular, non-lifestyle studies were ten times more often echoed when they reported initial findings than subsequent observations. Of these 102 primary studies, 53 (52.0%) were confirmed by the corresponding meta-analyses. The replication validity of the 28 subsequent non-lifestyle observations was still higher than that of the 29 initial findings (71.4% versus 41.4%; Chi2 test, $X^2 = 5.22$ $p = 0.022$). Likewise, studies related to psychiatry and covered by newspapers were more often initial findings than those in neurology (Fisher test; $p = 0.0072$) and for somatic diseases (Chi2 test $X^2 = 9.59$ $p = 0.002$). They were also confirmed less often (30.4%) than those related to neurological diseases (64%) (chi2 test; $X^2 = 5.41$ $p = 0.020$) and to somatic diseases (55.6) ($X^2 = 4.08$ $p = 0.043$).

Lack of media coverage of studies contradicting initial claims

Among the 53 initial studies covered by newspapers, 35 (66%) were disconfirmed by the corresponding meta-analyses. These 35 initial studies were followed by 503 subsequent studies of which 398 reported either the absence of a statistically significant effect or a significant effect in the opposite direction. Only one of these 398 subsequent studies and only one of the 35 corresponding meta-analysis articles were covered by newspapers. In the first case, the initial study published in 1999 by *The Lancet* (IF = 39) reported that the dopamine transporter level was increased (+70%) in both striatum of ADHD patients [26]. This initial study was covered by 22 newspaper articles. In contrast, according to a subsequent study published in 2009 by *JAMA* (IF = 30), the dopamine transporter level was significantly decreased (-19%) in the left striatum of ADHD patients, but not in the right [27]. The three newspapers that covered this subsequent study did not mention that it contradicted the initial claim although two newspapers (*The Globe and Mail* and *The Washington Post*) had also covered the initial study.

In the second case, the initial study was published in 2003 in *Science* and claimed that a genetic factor was associated with depression when subjects were exposed to stressful life events [28]. This finding was widely covered by newspapers (50 articles). Two subsequent studies that confirmed the same association were also echoed (7 and 2 newspaper articles). Newspapers never covered the eleven subsequent studies that failed to replicate this genetic association. Finally, the meta-analysis published in 2009 by Risch and coworkers in *JAMA* (IF = 30) and contradicting "positive" findings [29], was covered by only four newspaper articles and all four mentioned that the initial finding "has not held up to scientific scrutiny". Overall, 234 newspaper articles covered the 35 initial studies that were later disconfirmed by subsequent meta-analyses. However, only four newspaper articles explicitly mentioned the refutation of an initial claim previously covered in the press.

Comments

We explored the replication validity of association studies echoed by the lay press. Altogether, 48.7% of the studies examined were confirmed by subsequent meta-analyses. However, sub-population analyses show large variations of this replication rate. Newspapers preferentially echo initial studies over subsequent studies investigating the same question and initial findings covered by newspapers are less often validated by replications (34%) than subsequent findings (56.3%). This observation is consistent with several observational studies reporting that many initial or early biomedical findings are disconfirmed by subsequent studies [6–16]. Thus, the present study further supports our previous conclusion [19]: a major factor that explains the poor replication validity of biomedical studies covered by newspapers is that journalists favor initial findings over replication studies.

In agreement with previous studies [24, 25], we observed that newspapers preferentially reported lifestyle association studies linking a pathology to a risk factor on which each reader can act. Non-lifestyle studies related to brain imaging, genetic factor or other inescapable risk factors were less often echoed. This preferential coverage further supports the view that the first journalists' aim is to attract readers' attention. We also observed that newspapers never covered initial findings reporting null findings and under-reported subsequent null findings, particularly regarding non-lifestyle studies. As a consequence, journalists rarely cover null findings that contradict previously echoed studies. Within our database we observed only one interesting exception to this general rule: four newspaper articles covered the meta-analysis by Risch and coworkers and informed the public that it disconfirmed Caspi's initial study, which received much larger coverage (50 articles). All these observations support the conclusion of a case study: "scientific uncertainty *per se* is not attractive to journalists" (p. 344) [30].

We observed that findings reported by newspapers and related to psychiatric disorders were less often confirmed by meta-analyses than those related to neurological or somatic diseases. We correlated this observation with the fact that the preferential coverage of initial studies is even stronger regarding psychiatry. However, another factor might contribute to the low replication validity of psychiatric studies covered by newspapers: we previously showed that initial association studies linking biomarkers or risk factors to psychiatric disorders were less often confirmed by corresponding meta-analyses than those related to neurological diseases [11]. Likewise, highly cited early studies investigating the effectiveness of psychiatric treatments are often contradicted by subsequent studies [16].

Limitations

Although single primary studies have often investigated several associations linking a biomarkers or risk factors to a pathology (e.g. several polymorphisms on the same gene), we restricted our replication analysis to only one association per scientific article. Because we selected the most credible association (i.e. with the highest degree of statistical significance) it is likely that the replication rates we reported here are overestimated.

The 4723 primary articles of our database were included in meta-analyses thus enabling us to assess their replication validity. Whether the newspaper coverage of association studies included in meta-analyses is representative of the coverage of all association studies remains an open question.

We only considered newspaper articles published within a month following the publication of each scientific study. Indeed, when newspaper articles were published much later, it was often impossible to identify the scientific study they were reporting on with certainty. Consequently, we did not investigate the long-term impact of scientific publications in the lay press, although there is no doubt that certain scientific studies did have such an impact. For example, Caspi's study published in 2003 about a genetic susceptibility to major depression [28] was still cited in the lay press well after its refutation in 2009 by a meta-analysis [29]. However, in relative terms, it is likely that the scientific publications with the strongest long-term impact also received the widest newspaper coverage shortly following their publication.

Our data support three conclusions. First, newspapers preferentially cover lifestyle associations over non-lifestyle ones. Second, they strongly favor initial findings over subsequent studies when they cover non-lifestyle associations, but they equally cover initial and subsequent lifestyle studies. Third, among non-lifestyle associations covered by newspapers, initial findings are less often confirmed by meta-analyses than subsequent studies. However, our number of observations is too small to know whether this third conclusion also holds true regarding

lifestyle studies. Indeed, our database included only five initial lifestyle studies covered by newspapers.

We only studied newspaper articles but ignored television reporting, and this although television is the main source of information on medical sciences for Europeans (European Commission, 2007). Indeed, television programs rarely explicitly report on specific scientific publications [31]. Nevertheless, newspaper articles might indirectly influence the content of television programs by raising the attention of their producers.

Conclusion

Biomedical researches mature from early studies that are highly uncertain to scientific consensus built on subsequent studies and quantified by meta-analyses. Indeed, recent meta-researches have consistently shown that most early studies report inflated effect size or false positive effects [6–16]. Our study shows that many biomedical findings reported by newspapers are disconfirmed by subsequent studies. This is partly due to the fact that newspapers preferentially cover "positive" initial studies rather than subsequent observations, in particular those reporting null findings. Our study also suggests that most journalists from the general press do not know or prefer not to deal with the high degree of uncertainty inherent in early biomedical studies. Importantly, such biased newspaper coverage can have important social consequences [32, 33]. For example, a content analysis of newspaper articles covering Caspi's study showed that they emphasized the genetic side of the gene-by-environment interactions and this discourse deflected public attention away from the considerable impact of social inequalities upon health [34]. Therefore, with Susan Watts [35], we advocate that society "needs science journalism to weigh up the values and the vices of new science" (p. 151). In particular, when preparing a report on a scientific study, journalists should always ask scientists whether it is an initial finding and, if so, they should inform the public that this discovery is still tentative and must be validated by subsequent studies. Larsson and coworkers (2003) have identified the obstacles science journalists meet to accurately cope with uncertainty [36]. In particular, most interviewed journalists feel that it is difficult to find scientists who are independent from authors and who are willing to assist them. Thus, scientists, either as independent experts or as authors, are also responsible for improving the informative value of biomedical reporting in the mass media. In particular, they are responsible for the accuracy of the press releases covering their work and published by scientific editors or universities. This is not only a moral duty but also in the interest of science. Indeed, contrary to common beliefs, reception studies show that scientists are viewed by the public as more trustworthy when news coverage of health research acknowledges its uncertainty, especially if this is attributed to scientific authors [37, 38].

Supporting information

S1 Text. References of the 156 studies reported by newspapers and of their corresponding meta-analysis articles.

(DOCX)

S2 Text. Raw data shown in Figs 1–3 and about primary studies covered by three newspapers or more.

(DOCX)

S1 File. Raw data about the 156 studies covered by newspapers and their corresponding meta-analyses.

(XLSX)

Author Contributions

Conceptualization: EDM AS TB FG.

Data curation: EDM TB FG.

Formal analysis: EDM TB FG.

Investigation: EDM FG.

Methodology: EDM FG.

Project administration: EDM AS TB FG.

Supervision: AS TB FG.

Validation: FG.

Visualization: EDM TB FG.

Writing – original draft: EDM FG.

Writing – review & editing: EDM AS TB FG.

References

- Ioannidis JP. Why most published research findings are false. *PLoS Med.* 2005; 2(8):e124. doi: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124) PMID: [16060722](https://pubmed.ncbi.nlm.nih.gov/16060722/)
- Sarewitz D. Beware the creeping cracks of bias. *Nature.* 2012; 485(7397):149. doi: [10.1038/485149a](https://doi.org/10.1038/485149a) PMID: [22575922](https://pubmed.ncbi.nlm.nih.gov/22575922/)
- Munafo MR, Stothart G, Flint J. Bias in genetic association studies and impact factor. *Mol Psychiatry.* 2009; 14(2):119–20. doi: [10.1038/mp.2008.77](https://doi.org/10.1038/mp.2008.77) PMID: [19156153](https://pubmed.ncbi.nlm.nih.gov/19156153/)
- Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res.* 2015; 116(1):116–26. doi: [10.1161/CIRCRESAHA.114.303819](https://doi.org/10.1161/CIRCRESAHA.114.303819) PMID: [25552691](https://pubmed.ncbi.nlm.nih.gov/25552691/)
- Baker M. 1,500 scientists lift the lid on reproducibility. *Nature.* 2016; 533(7604):452–4. doi: [10.1038/533452a](https://doi.org/10.1038/533452a) PMID: [27225100](https://pubmed.ncbi.nlm.nih.gov/27225100/)
- Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov.* 2011; 10(9):712. doi: [10.1038/nrd3439-c1](https://doi.org/10.1038/nrd3439-c1) PMID: [21892149](https://pubmed.ncbi.nlm.nih.gov/21892149/)
- Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. *PLoS Biol.* 2015; 13(6):e1002165. doi: [10.1371/journal.pbio.1002165](https://doi.org/10.1371/journal.pbio.1002165) PMID: [26057340](https://pubmed.ncbi.nlm.nih.gov/26057340/)
- Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet.* 2001; 29(3):306–9. doi: [10.1038/ng749](https://doi.org/10.1038/ng749) PMID: [11600885](https://pubmed.ncbi.nlm.nih.gov/11600885/)
- Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG, Ioannidis JP. Establishment of genetic associations for complex diseases is independent of early study findings. *Eur J Hum Genet.* 2004; 12(9):762–9. doi: [10.1038/sj.ejhg.5201227](https://doi.org/10.1038/sj.ejhg.5201227) PMID: [15213707](https://pubmed.ncbi.nlm.nih.gov/15213707/)
- Ioannidis JP, Panagiotou OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA.* 2011; 305(21):2200–10. doi: [10.1001/jama.2011.713](https://doi.org/10.1001/jama.2011.713) PMID: [21632484](https://pubmed.ncbi.nlm.nih.gov/21632484/)
- Dumas-Mallet E, Button K, Boraud T, Munafo M, Gonon F. Replication Validity of Initial Association Studies: A Comparison between Psychiatry, Neurology and Four Somatic Diseases. *PLoS One.* 2016; 11(6):e0158064. doi: [10.1371/journal.pone.0158064](https://doi.org/10.1371/journal.pone.0158064) PMID: [27336301](https://pubmed.ncbi.nlm.nih.gov/27336301/)
- Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. *Proc Natl Acad Sci U S A.* 2001; 98(3):831–6. doi: [10.1073/pnas.98.3.831](https://doi.org/10.1073/pnas.98.3.831) PMID: [11158556](https://pubmed.ncbi.nlm.nih.gov/11158556/)
- Trikalinos TA, Churchill R, Ferri M, Leucht S, Tuunainen A, Wahlbeck K, et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol.* 2004; 57(11):1124–30. PMID: [15612138](https://pubmed.ncbi.nlm.nih.gov/15612138/)
- Gan HK, You B, Pond GR, Chen EX. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. *J Natl Cancer Inst.* 2012; 104(8):590–8. doi: [10.1093/jnci/djs141](https://doi.org/10.1093/jnci/djs141) PMID: [22491345](https://pubmed.ncbi.nlm.nih.gov/22491345/)

15. Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. *JAMA*. 2012; 308(16):1676–84. doi: [10.1001/jama.2012.13444](https://doi.org/10.1001/jama.2012.13444) PMID: [23093165](https://pubmed.ncbi.nlm.nih.gov/23093165/)
16. Tajika A, Ogawa Y, Takeshima N, Hayasaka Y, Furukawa TA. Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up. *Br J Psychiatry*. 2015; 207(4):357–62. doi: [10.1192/bjp.bp.113.143701](https://doi.org/10.1192/bjp.bp.113.143701) PMID: [26159600](https://pubmed.ncbi.nlm.nih.gov/26159600/)
17. Open-Science-Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science*. 2015; 349(6251):aac4716. doi: [10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716) PMID: [26315443](https://pubmed.ncbi.nlm.nih.gov/26315443/)
18. Bacchetti P, Deeks SG, McCune JM. Breaking free of sample size dogma to perform innovative translational research. *Sci Transl Med*. 2011; 3(87):87ps24. doi: [10.1126/scitranslmed.3001628](https://doi.org/10.1126/scitranslmed.3001628) PMID: [21677197](https://pubmed.ncbi.nlm.nih.gov/21677197/)
19. Gonon F, Konsman JP, Cohen D, Boraud T. Why most biomedical findings echoed by newspapers turn out to be false: the case of Attention Deficit Hyperactivity Disorder. *PLoS One*. 2012; 7(9):e44275. doi: [10.1371/journal.pone.0044275](https://doi.org/10.1371/journal.pone.0044275) PMID: [22984483](https://pubmed.ncbi.nlm.nih.gov/22984483/)
20. Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry*. 2006; 163(2):185–94. doi: [10.1176/appi.ajp.163.2.185](https://doi.org/10.1176/appi.ajp.163.2.185) PMID: [16449469](https://pubmed.ncbi.nlm.nih.gov/16449469/)
21. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008; 358(3):252–60. doi: [10.1056/NEJMsa065779](https://doi.org/10.1056/NEJMsa065779) PMID: [18199864](https://pubmed.ncbi.nlm.nih.gov/18199864/)
22. Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, et al. Coverage by the news media of the benefits and risks of medications. *N Engl J Med*. 2000; 342(22):1645–50. doi: [10.1056/NEJM200006013422206](https://doi.org/10.1056/NEJM200006013422206) PMID: [10833211](https://pubmed.ncbi.nlm.nih.gov/10833211/)
23. Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, Mantz J, et al. Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. *PLoS Med*. 2012; 9(9):e1001308. doi: [10.1371/journal.pmed.1001308](https://doi.org/10.1371/journal.pmed.1001308) PMID: [22984354](https://pubmed.ncbi.nlm.nih.gov/22984354/)
24. Bartlett C, Sterne J, Egger M. What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers. *BMJ*. 2002; 325(7355):81–4. PMID: [12114239](https://pubmed.ncbi.nlm.nih.gov/12114239/)
25. Stryker JE. Reporting medical information: effects of press releases and newsworthiness on medical journal articles' visibility in the news media. *Prev Med*. 2002; 35(5):519–30. PMID: [12431901](https://pubmed.ncbi.nlm.nih.gov/12431901/)
26. Dougherty DD, Bonab AA, Spencer TJ, Rauch SL, Madras BK, Fischman AJ. Dopamine transporter density in patients with attention deficit hyperactivity disorder. *Lancet*. 1999; 354(9196):2132–3. doi: [10.1016/S0140-6736\(99\)04030-1](https://doi.org/10.1016/S0140-6736(99)04030-1) PMID: [10609822](https://pubmed.ncbi.nlm.nih.gov/10609822/)
27. Volkow ND, Wang GJ, Kollins SH, Wigal TL, Newcorn JH, Telang F, et al. Evaluating dopamine reward pathway in ADHD: clinical implications. *JAMA*. 2009; 302(10):1084–91. doi: [10.1001/jama.2009.1308](https://doi.org/10.1001/jama.2009.1308) PMID: [19738093](https://pubmed.ncbi.nlm.nih.gov/19738093/)
28. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*. 2003; 301(5631):386–9. doi: [10.1126/science.1083968](https://doi.org/10.1126/science.1083968) PMID: [12869766](https://pubmed.ncbi.nlm.nih.gov/12869766/)
29. Risch N, Herrell R, Lehner T, Liang KY, Eaves L, Hoh J, et al. Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA*. 2009; 301(23):2462–71. doi: [10.1001/jama.2009.878](https://doi.org/10.1001/jama.2009.878) PMID: [19531786](https://pubmed.ncbi.nlm.nih.gov/19531786/)
30. Kitzinger J, Reilly J. The rise and fall of risk reporting: media coverage of human genetics research, 'false memory syndrome' and 'mad cow disease'. *Eur J Communication*. 1997; 12(3):319–50.
31. Bourdaa M, Konsman JP, Secail C, Venturini T, Veyrat-Masson I, Gonon F. Does television reflect the evolution of scientific knowledge? The case of attention deficit hyperactivity disorder coverage on French TV. *Public Understanding of Science*. 2015; 24(2):200–9. doi: [10.1177/0963662513484842](https://doi.org/10.1177/0963662513484842) PMID: [23825292](https://pubmed.ncbi.nlm.nih.gov/23825292/)
32. Holtzman NA, Bernhardt BA, Mountcastle-Shah E, Rodgers JE, Tambor E, Geller G. The quality of media reports on discoveries related to human genetic diseases. *Community Genet*. 2005; 8(3):133–44. doi: [10.1159/000086756](https://doi.org/10.1159/000086756) PMID: [16113530](https://pubmed.ncbi.nlm.nih.gov/16113530/)
33. Gonon F, Bezard E, Boraud T. What should be said to the lay public regarding ADHD etiology. *Am J Med Genet B Neuropsychiatr Genet*. 2011; 156(8):989–91.
34. Horwitz AV. Media portrayals and health inequalities: a case study of characterizations of Gene x Environment interactions. *J Gerontol B Psychol Sci Soc Sci*. 2005; 60 Spec No 2:48–52.
35. Watts S. Society needs more than wonder to respect science. *Nature*. 2014; 508(7495):151. doi: [10.1038/508151a](https://doi.org/10.1038/508151a) PMID: [24717473](https://pubmed.ncbi.nlm.nih.gov/24717473/)

36. Larsson A, Oxman AD, Carling C, Herrin J. Medical messages in the media—barriers and solutions to improving medical journalism. *Health Expect*. 2003; 6(4):323–31. doi: [10.1046/j.1369-7625.2003.00228.x](https://doi.org/10.1046/j.1369-7625.2003.00228.x) PMID: [15040794](https://pubmed.ncbi.nlm.nih.gov/15040794/)
37. Jensen JD, Hurley RJ. Conflicting stories about public scientific controversies: Effects of news convergence and divergence on scientists' credibility. *Public Underst Sci*. 2012; 21(6):689–704. doi: [10.1177/0963662510387759](https://doi.org/10.1177/0963662510387759) PMID: [23832155](https://pubmed.ncbi.nlm.nih.gov/23832155/)
38. Jensen JD, Krakow M, John KK, Liu M. Against conventional wisdom: when the public, the media, and medical practice collide. *BMC Med Inform Decis Mak*. 2013; 13 Suppl 3:S4.

Tables SI-S3 give the raw data shown in percentage in Figures 1-3. These tables and figures only take into account primary studies; meta-analyses articles were not considered.

Table S1: raw data of Figure 1

Study type	number of primary studies in the database				number of covered studies			
	all	0 < IF ≤ 10	10 < IF < 30	IF ≥ 30	all	0 < IF ≤ 10	10 < IF < 30	IF ≥ 30
all studies								
initial	405	287	72	46	53	8	15	30
subsequent	4318	3887	379	52	103	44	32	27
all	4723	4174	451	98	156	52	47	57
lifestyle								
initial	39	30	5	4	5	1	1	3
subsequent	600	538	49	13	58	27	21	10
all	639	568	54	17	63	28	22	13
non-lifestyle								
initial	366	257	67	42	48	7	14	27
subsequent	3718	3349	330	39	45	17	11	17
all	4084	3606	397	81	93	24	25	44

Table S2: raw data of Figure 2

Study type	number of covered studies				number of covered and confirmed			
	all	0 < IF ≤ 10	10 < IF < 30	IF ≥ 30	all	0 < IF ≤ 10	10 < IF < 30	IF ≥ 30
all studies								
initial	53	8	15	30	18	2	4	12
subsequent	103	44	32	27	58	20	17	21
all	156	52	47	57	76	22	21	33
lifestyle								
initial	5	1	1	3	2	0	1	1
subsequent	58	27	21	10	29	14	9	6
all	63	28	22	13	31	14	10	7
non-lifestyle								
initial	48	7	14	27	16	2	3	11
subsequent	45	17	11	17	29	6	8	15
all	93	24	25	44	45	8	11	26

Table S3: raw data of Figure 3

Study type	number of studies covered			number of covered and confirmed		
	PSY	NEURO	SOMA	PSY	NEURO	SOMA
initial	23	11	19	5	4	9
subsequent	15	30	58	5	22	31
all	38	41	77	10	26	40

Subpopulation analysis of primary studies covered by three newspapers or more.

Table S4: Preferential coverage of initial findings and influence of the impact factor (IF).

Study type	number of primary studies in the database				number of covered studies			
	all	0 < IF ≤ 10	10 < IF < 30	IF ≥ 30	all	0 < IF ≤ 10	10 < IF < 30	IF ≥ 30
all studies								
initial	405	287	72	46	32	4	5	23
subsequent	4318	3887	379	52	70	23	23	24
all	4723	4174	451	98	102	27	28	47
lifestyle								
initial	39	30	5	4	3	1	0	2
subsequent	600	538	49	13	42	15	17	10
all	639	568	54	17	45	16	17	12
non-lifestyle								
initial	366	257	67	42	29	3	5	21
subsequent	3718	3349	330	39	28	8	6	14
all	4084	3606	397	81	57	11	11	35

Table S5: Replication validity of primary articles reported by three newspapers or more.

Study type	number of covered studies				number of covered and confirmed			
	all	0 < IF ≤ 10	10 < IF < 30	IF ≥ 30	all	0 < IF ≤ 10	10 < IF < 30	IF ≥ 30
all studies								
initial	32	4	5	23	13	1	1	11
subsequent	70	23	23	24	40	8	13	19
all	102	27	28	47	53	9	14	30
lifestyle								
initial	3	1	0	2	1	0	0	1
subsequent	42	15	17	10	20	6	8	6
all	45	16	17	12	21	6	8	7
non-lifestyle								
initial	29	3	5	21	12	1	1	10
subsequent	28	8	6	14	20	2	5	13
all	57	11	11	35	32	3	6	23

Table S6: Replication validity of primary studies reported by three newspapers or more in three biomedical domains.

Study type	number of studies covered			number of covered and confirmed		
	PSY	NEURO	SOMA	PSY	NEURO	SOMA
initial	14	5	13	4	2	7
subsequent	9	20	41	3	14	23
all	23	25	54	7	16	30

ADHD

initial 4

Marker	# of lines 6						individual study										Meta-analysis									
	ES	path	type	S	order	ini	Year	Author	Sign	IF	media	ES	CI low	CI high	p	[LnES]	Year	Author	Sign	ES	CI low	CI high	p	[LnES]	consist	replic
DRD2-TaqI rs1800497 allelic	OR	ADHD	G	0	ini	1	1991	Comings	Y	30	9	4,7	2,26	9,78	0,0001	1,5475625	2012	Wu	Y	1,65	1,05	2,58	0,029	0,5007753	YY	0
DAT 10-R 3'UTR VNTR allelic	OR	ADHD	G	0	ini	1	1995	Cook	Y	11,2	3	2,84	1,3243	6,0995	0,0073	1,0436041	2009	Gizer	Y	1,1	1	1,27	0,028	0,0953102	YY	0
DAT1 imaging	d	ADHD	I	0	ini	1	1999	Dougherty	Y	39	22	2,37	1,35	3,39	0,0001	4,298706	2012	Fusar-Poli	Y	0,23	0,01	0,46	0,04	0,417174	YY	0
DRD4 gene 7R carrier	OR	ADHD	G	0	ini	1	1996	LaHoste	Y	14,9	9	3,01	1,29	7,05	0,011	1,1019401	2010	Smith	Y	1,33	1,16	1,53	0,0001	0,2851789	YY	0
Colorings (parents)	d	ADHD	S	1	op	0	2007	McCann	Y	39	47	0,198			0,044	0,3591324	2012	Nigg	Y	0,18	0,08	0,29	0,0007	0,326484	YY	1
DAT1 imaging	d	ADHD	I	0	op	0	2009	Volkow	Y	30	3	-0,62	-1,02	-0,21	0,003	1,124556	2012	Fusar-Poli	Y	0,23	0,01	0,46	0,04	0,417174	YY	0

Comments

column B	indicates how effect size were expressed: OR: odd ratio, d: standardized mean difference
column D	type of study: G: genetic, I: brain imaging, S: risk factor on which each subject can act, O: other
column F	ini: initial study, op: other primary study, meta: meta-analysis
column I	first author
column J	indicates whether the individual study reported a statistically significant effect. Y: yes, N: no
column K	IF: 2012 impact factor of the journal that published the individual study
column L	number of newspaper articles echoing each individual study
column M	ES: effect size
columns N and O	CI: confidence interval
column P	statistical significance (p value)
column Q	logarithm of the effect size
column T	first author of the meta-analysis corresponding to each individual study
column U	indicates whether the corresponding meta-analysis reported a statistically significant effect. Y: yes, N: no
column V	ES: effect size
columns W and X	CI: confidence interval
column Y	statistical significance (p value)
column Z	logarithm of the effect size
column AA	consistency: nominal agreement between each individual study and the corresponding meta-analysis (YY: the individual study and the meta-analysis reported significant effects in the opposite direction)
column AB	replication validity: the individual study must be in nominal agreement with the corresponding meta-analysis and its effect size (in log) must not be inflated by more than 100%
marker	indicates that this individual study has been published after another individual study on the same topic and which has been also covered by newspapers

AUTISM

initial 4

Marker	# lignes				individual study										Meta-analysis											
	ES	path	type	S	order	ini	Year	Author	Sign	IF	media	ES	CI low	CI high	p	[LnES]	Year	Author	Sign	ES	CI low	CI high	p	[LnES]	consist	replic
BHTTLPR S/L family-based trios	OR	AUT	G	0	ini	1	1997	Cook	Y	14.9	1	1.66	1.04	2.62	0.0315	0.5068176	2008	Huang	N	1.03	0.842	1.265		0.0295588	YN	0
Vermal lobules VI-VII (adult)	d	AUT	I	0	ini	1	1998	Courchesne	Y	51.7	8	-1.22	-2.02	-0.431	0.0026	2.212336	2008	Stanfield	N	-0.12	-0.36	0.13	0.35	0.217656	YN	0
Paternal age 40-49 / <29	RR	AUT	S	1	ini	1	2004	Glasson	Y	13.8	2	1.58	1.0847	2.299	0.0171	0.4574248	2011	Hultman	YY	1.78	1.52	2.07	0.0001	0.57661336	YY	1
Homocysteine plasma/serum	d	AUT	O	0	ini	1	2004	James	Y	6.5	2	-0.49			0.01	0.888762	2012	Frustaci	N	-0.09	-0.587	0.389	0.691	0.163242	YN	0
Paternal age 40-49 / <29	RR	AUT	S	1	op	0	2006	Reichenberg	Y	13.8	40	5.75	2.65	12.46	0.0001	1.7491999	2011	Hultman	YY	1.78	1.52	2.07	0.0001	0.57661336	YY	0
Total Brain Volume (children)	d	AUT	I	0	op	0	2002	Sparks	Y	8.2	1	1.03	0.463	1.589	0.0003	1.868214	2008	Stanfield	YY	0.41	0.18	0.63	0.0004	0.743658	YY	0

Comments

- column B indicates how effect size were expressed: OR: odd ratio, d: standardized mean difference
- column D type of study: G: genetic, I: brain imaging, S: risk factor on which each subject can act, O: other
- column F ini: initial study, op: other primary study, meta: meta-analysis
- column I first author
- column J indicates whether the individual study reported a statistically significant effect. Y: yes, N: no
- column K IF: 2012 impact factor of the journal that published the individual study
- column L number of newspaper articles echoing each individual study
- column M ES: effect size
- columns N and O CI: confidence interval
- column P statistical significance (p value)
- column Q logarithm of the effect size
- column T first author of the meta-analysis corresponding to each individual study
- column U indicates whether the corresponding meta-analysis reported a statistically significant effect. Y: yes, N: no
- column V ES: effect size
- columns W and X CI: confidence interval
- column Y statistical significance (p value)
- column Z logarithm of the effect size
- column AA consistency: nominal agreement between each individual study and the corresponding meta-analysis (YYO: the individual study and the meta-analysis reported significant effects in the opposite direction)
- column AB replication validity: the individual study must be in nominal agreement with the corresponding meta-analysis and its effect size (in log) must not be inflated by more than 100%
- marker indicates that this individual study has been published after another individual study on the same topic and which has been also echoed by newspapers

MDD

initial 4

Marker	# of lines				individual study										Meta-analysis											
	ES	path	type	S	order	ini	Year	Author	Sign	IF	media	ES	CI low	CI high	p	[LnES]	Year	Author	Sign	ES	CI low	CI high	p	[LnES]	consist	repro
homocysteine : high plasma level	OR	MDD	O	0	op	0	2008	Almeida	Y	13.8	1	1.53	1.1	2.11	0.011	0.42526774	2008	Almeida	YY	1.7	1.38	2.08	0.0001	0.5306283	YY	1
5-HTTLPR S/L risk x stress	OR	MDD	G	0	ini	1	2003	Caspi	Y	31	50	1.3	1.06	1.67	0.024	0.26236426	2009	Rish	NN	1.01	0.94	1.1		0.009593	YN	0
5-HTTLPR S/L risk SS/SL+LL	OR	MDD	G	0	ini	1	1996	Collier	Y	14.9	2	1.75	1.15	2.62	0.008	0.55961579	2010	Clarke	YY	1.16	1.08	1.24	0.0001	0.14842	YY	0
Low birth weight	OR	MDD	O	0	op	0	2007	Colman	Y	9.2	5	1.15	1.08	1.22	0.0001	0.13976194	2013	Wojcik	Y	1.15	1	1.32	0.0485	0.1397619	YY	1
SGPFC left volume	d	MDD	I	0	ini	1	1997	Drevets	Y	38.6	5	-1.08	-1.76	-0.4	0.0019	1.958904	2012	Arnone	N	-0.37	-0.82	0.08	0.05	0.671106	YN	0
Low birth weight	OR	MDD	O	0	op	0	2004	Gale	Y	6.6	1	1.5	1.1	2	0.008	0.40546511	2013	Wojcik	Y	1.15	1	1.32	0.0485	0.1397619	YY	0
5-HTTLPR x amygdala acti	d	MDD	I	0	ini	1	2002	Hariri	Y	31	18	1.04	0.25	1.83	0.01	1.886352	2008	Munafò	YY	0.55	0.353	0.747	0.0001	0.99759	YY	1
Low birth weight	OR	MDD	O	0	op	0	2004	Patton	Y	6.6	1	11.6	2.2	62	0.004	2.4510051	2013	Wojcik	Y	1.15	1	1.32	0.0485	0.1397619	YY	0
5-HTTLPR S/L risk x stress	OR	MDD	G	0	op	0	2006	Taylor	Y	9.2	2	1.6			0.006	0.47303363	2009	Rish	NN	1.01	0.94	1.1		0.009593	YN	0
Low birth weight	OR	MDD	O	0	op	0	2005	Wiles	Y	6.6	3	1.49	1.01	2.2	0.045	0.38877612	2013	Wojcik	Y	1.15	1	1.32	0.0485	0.1397619	YY	0
5-HTTLPR S/L risk x stress	OR	MDD	G	0	op	0	2006	Wilhelm	Y	6.6	7	1.7	1.1	3	0.036	0.53062825	2009	Rish	NN	1.01	0.94	1.1		0.009593	YN	0

Meta-analysis echoed by newspapers

Marker	ES	path	type	S	order	ini	Year	Author	Sign	IF	media	ES	CI low	CI high	p	[LnES]
5-HTTLPR S/L risk x stress	OR	MDD	G	0	meta		2009	Rish	N	30	4	1.01	0.94	1.1		0.0095933
Postpartum 3 to 6 months	d	MDD	O	0	meta		2010	Paulson	Y	30	41	0.26	0.17	0.36	0.0001	1.34707365

Comments

- column B indicates how effect size were expressed: OR: odd ratio, d: standardized mean difference
- column D type of study: G: genetic, I: brain imaging, S: risk factor on which each subject can act, O: other
- column F ini: initial study, op: other primary study, meta: meta-analysis
- column I first author
- column J indicates whether the individual study reported a statistically significant effect. Y: yes, N: no
- column K IF: 2012 impact factor of the journal that published the individual study
- column L number of newspaper articles echoing each individual study
- column M ES: effect size
- columns N and O CI: confidence interval
- column P statistical significance (p value)
- column Q logarithm of the effect size
- column T first author of the meta-analysis corresponding to each individual study
- column U indicates whether the corresponding meta-analysis reported a statistically significant effect. Y: yes, N: no
- column V ES: effect size
- columns W and X CI: confidence interval
- column Y statistical significance (p value)
- column Z logarithm of the effect size
- column AA consistency: nominal agreement between each individual study and the corresponding meta-analysis (YO: the individual study and the meta-analysis reported significant effects in the opposite direction)
- column AB replication validity: the individual study must be in nominal agreement with the corresponding meta-analysis and its effect size (in log) must not be inflated by more than 100%
- marker indicates that this individual study has been published after another individual study on the same topic and which has been also echoed by newspapers

Schizophrenia

ini 11

Marker	# of lines		Individual study																Meta-analysis									
	ES	path	type	S	order	ini	Year	Author	Sign	IF	media	ES	CI low	CI high	p	[LNEs]	Year	Author	Sign	ES	CI low	CI high	p	[LNEs]	consist	replic		
CR22 Ser311Cys rs1801028 G vs C asian	OR	SCH	G	0	ini	1	1994	Akinami	Y	39	1	3.09	1.43	6.67	0.0041	1.1281711	2012	Liu	YY	1.47	1.18	1.83	0.0006	0.3852624	YY	0		
Paternal age 250 / 25-29	RR	SCH	S	1	op	0	2003	Byrne	Y	13.8	1	1.75	1.41	2.16		0.55961579	2011	Miller	YY	1.66	1.46	1.89	0.0001	0.5068176	YY	1		
G72/G30 rs78293 M22 allelic	OR	SCH	G	0	ini	1	2002	Chumakov	Y	9.7	1	1.51				0.003	0.4121097	2008	Shi	N	0.99	0.87	1.13	0.89	0.01005034	YN	0	
Left lateral ventricles FES (MRI)	d	SCH	I	0	op	0	2000	Fannon	Y	14.7	3	0.62	0.1	1.14	0.019	1.124556	2012	De Peri	YY	0.49	0.35	0.64	0.0001	0.888762	YY	1		
PPP3CC rs2461491 A/T allelic CC-S3	OR	SCH	G	0	ini	1	2003	Gerber	Y	9.7	3	1.15				0.041	0.1367619	2008	Shi b	YY	1.07	1.022	1.239	0.004	0.06763865	YY	0	
CAFBR2 rs253944 C/T allelic	OR	SCH	G	0	ini	1	2004	Lo	Y	14.9	1	2.34	1.46	3.75	0.0003	0.8501500	2008	Shi b	N	1.06	0.862	1.290	0.569	0.05826891	YN	0		
Paternal age 250 / 25-29	RR	SCH	S	1	op	0	2001	Malaspina	Y	13.8	29	2.96	1.6	5.47		1.08518927	2011	Miller	YY	1.66	1.46	1.89	0.0001	0.5068176	YY	0		
ZNF804A rs1344706 G/T allelic	OR	SCH	G	0	ini	1	2008	O'Donovan	Y	35.2	14	1.14	1.04	1.26	0.0074	0.1310283	2012	Zhang	YY	1.12	1.06	1.17	0.0001	0.11332869	YY	1		
Superior temporal gyrus GM MRI	d	SCH	I	0	ini	1	1992	Shenlon	Y	51.7	8	-1.01	-1.72	-0.29	0.002	1.831938	2013	Hajima	YY	-0.577	-0.75	-0.41	0.0001	1.0465626	YY	1		
COMT rs155559 allelic	OR	SCH	G	0	ini	1	2002	Shifman	Y	11.2	3	1.24	1.106	1.38	0.0002	0.2151114	2009	Koochi	NN	1.03	0.96	1.108	0.396	0.0295588	YN	0		
Paternal age 250 / 25-29	RR	SCH	S	1	op	0	2004	Sipos	Y	17.2	12	5.85	2.91	11.74		1.76644166	2011	Miller	YY	1.66	1.46	1.89	0.0001	0.5068176	YY	0		
NRG1 SNP8NRG221533 rs35753505 al	OR	SCH	G	0	ini	1	2002	Stefansson	Y	11.2	1	1.35	1.09	1.68	0.0057	0.3001046	2008	Munaf0	NN	1.04	0.99	1.1	0.1	0.03922071	YN	0		
DTNBP1 rs3213207 G/C allelic	OR	SCH	G	0	ini	1	2002	Straub	Y	11.2	4	0.62	0.41	0.93	0.0001	0.4780358	2008	Shi b	Y	0.91	0.847	0.989	0.025	0.09431068	YY	0		
Hippocampus volume MRI	d	SCH	I	0	ini	1	1990	Suddath	Y	51.7	7	-0.6				0.01	1.08828	2013	Hajima	YY	-0.52	-0.6	-0.44	0.0001	0.9431176	YY	1	
CH3L1 rs4950628 G/C allelic	OR	SCH	G	0	ini	1	2007	Zhao	Y	11.2	1	1.49				0.004	0.3987761	2010	Chi	N	1.03	0.86	1.24	0.72	0.0295588	YN	0	

Comments

- column B indicates how effect size were expressed: OR: odd ratio, d: standardized mean difference
- column D type of study: G: genetic, I: brain imaging, S: risk factor on which each subject can act, O: other
- column F ini: initial study, op: other primary study, meta: meta-analysis
- column I first author
- column J indicates whether the individual study reported a statistically significant effect. Y: yes, N: no
- column K IF: 2012 impact factor of the journal that published the individual study
- column L number of newspaper articles echoing each individual study
- column M ES: effect size
- columns N and O CI: confidence interval
- column P statistical significance (p value)
- column Q logarithm of the effect size
- column T first author of the meta-analysis corresponding to each individual study
- column U indicates whether the corresponding meta-analysis reported a statistically significant effect. Y: yes, N: no
- column V ES: effect size
- columns W and X CI: confidence interval
- column Y statistical significance (p value)
- column Z logarithm of the effect size
- column AA consistency: nominal agreement between each individual study and the corresponding meta-analysis (YYO: the individual study and the meta-analysis reported significant effects in the opposite direction)
- column AB replication validity: the individual study must be in nominal agreement with the corresponding meta-analysis and its effect size (in log) must not be inflated by more than 100%
- marker indicates that this individual study has been published after another individual study on the same topic and which has been also echoed by newspapers

Alzheimer

ini 7

Marker	# of lines		0		individual study							Meta-analysis														
	ES	path	type	S	order	ini	Year	Author	Sign	IF	media	ES	CI low	CI high	p	[LnES]	Year	Author	Sign	ES	CI low	CI high	p	[LnES]	consist	replic
APCE 491A/T AA vs TT+AT rs449647	OR	AD	G	0	ini	1	1998	Bullido	Y	35.2	3	2.5	1.695	3.6634	0.0001	0.91629	2010	Xin	YY	1.49	1.29	1.72	0.0001	0.3265576	YY	0
Homocystein levels	G	AD	O	0	ini	1	1998	Clarke	Y	7.7	2	0.55	0.25	0.85	0.0003	0.98759	2011	Ho	YY	0.59	0.37	0.8	0.0001	1.070142	YY	1
Antihypertensive medication	OR	AD	O	0	ini	1	1998	Forette	Y	39	2	0.5			0.049	0.69315	2011	Chang-Quan	N	0.9	0.79	1.03	0.13	0.1053605	YN	0
PGBD1 rs3800324 allelic	OR	AD	G	0	ini	1	2007	Grupe (Wu)	Y	7.7	2	1.78	1.04	3.12	0.0396	0.57661	2011	Belbin	N	1.07	0.92	1.24	0.36	0.0676586	YN	0
CR1 rs3818361 allelic	OR	AD	G	0	op	0	2009	Harold UK	Y	35.2	46	1.18	1.079	1.293	0.0003	0.16551	2011	Antunez	YY	1.18	1.113	1.252	0.0001	0.1655144	YY	1
Antihypertensive medication	OR	AD	O	0	op	0	2006	Khachaturian	Y	7.7	7	0.64	0.41	0.98	0.045	0.44629	2011	Chang-Quan	N	0.9	0.79	1.03	0.13	0.1053605	YN	0
CR1 rs3818361 allelic	OR	AD	G	0	op	0	2009	Lambert	Y	35.2	36	1.19	1.11	1.26	0.0001	0.17395	2011	Antunez	YY	1.18	1.113	1.252	0.0001	0.1655144	YY	1
CY146A1 T/C TTTC+CC	OR	AD	G	0	op	0	2003	Papassotiroopoulos	Y	7.7	11	2.16	1.41	3.32	0.0004	0.77011	2013	Li b	NN	1.01	0.88	1.16		0.0099503	YN	0
Antihypertensive medication	OR	AD	O	0	op	0	2008	Peters	Y	23.9	8	0.87	0.76	1	0.045	0.13926	2011	Chang-Quan	N	0.9	0.79	1.03	0.13	0.1053605	YN	0
GAB2 rs10793294 allelic	OR	AD	G	0	ini	1	2007	Reiman NDC	Y	15.8	2	0.55	0.42	0.7	0.0001	0.59784	2011	Belbin	Y	0.78	0.64	0.93	0.007	0.2484614	YY	0
SORL1 A/G rs2282649 allelic	OR	AD	G	0	ini	1	2007	Rogaeva	Y	35.2	36	1.31	1.08	1.59	0.0062	0.27003	2010	Reynolds	YY	1.1	1.04	1.16	0.001	0.0953102	YY	0
APCE e4 late onset AD allelic	OR	AD	G	0	ini	1	1993	Strittmatter	Y	9.7	5	5.7	3.4391	9.3533	0.0001	1.74047	2012	Sadigh-Eleghad	YY	3.98	3.44	4.61	0.0001	1.3812818	YY	1

Comments

- column B indicates how effect size were expressed: OR: odd ratio, d: standardized mean difference
- column D type of study; G: genetic, I: brain imaging, S: risk factor on which each subject can act, O: other
- column F ini: initial study, op: other primary study, meta: meta-analysis
- column I first author
- column J indicates whether the individual study reported a statistically significant effect. Y: yes, N: no
- column K IF: 2012 impact factor of the journal that published the individual study
- column L number of newspaper articles echoing each individual study
- column M ES: effect size
- columns N and O CI: confidence interval
- column P statistical significance (p value)
- column Q logarithm of the effect size
- column T first author of the meta-analysis corresponding to each individual study
- column U indicates whether the corresponding meta-analysis reported a statistically significant effect. Y: yes, N: no
- column V ES: effect size
- columns W and X CI: confidence interval
- column Y statistical significance (p value)
- column Z logarithm of the effect size
- column AA consistency: nominal agreement between each individual study and the corresponding meta-analysis (YYO: the individual study and the meta-analysis reported significant effects in the opposite direction)
- column AB replication validity: the individual study must be in nominal agreement with the corresponding meta-analysis and its effect size (ln log) must not be inflated by more than 100%
- marker indicates that this individual study has been published after another individual study on the same topic and which has been also echoed by newspapers

Multiple sclerosis

ini 2

Marker	# of lines	Individual study					Meta-analysis																			
		ES	path	type	S	order	Year	Author	Sign	IF	media	ES	CI low	CI high	p	[LnES]	Year	Author	Sign	ES	CI low	CI high	p	[LnES]	consist	replic
Anti-VCA IgG (EBV)	OR	MS	O	0	op	0	2004	Alotata	Y	30	7	3,8	1,38	10,56	0,01	1,33500107	2010	Santiago	YY	5,5	3,37	8,81	0,0001	1,70474809	YY	1
KIF1B rs10492972 C/T allelic	OR	MS	G	0	op	0	2009	ANZgene	N	35,2	2	1	0,9169	1,0921	0,988	0	2011	Kudryavtseva	NN	1,02	0,97	1,06	0,46	0,01980263	NN	0
Anti-VCA IgG (EBV)	OR	MS	O	0	op	0	2001	Ascherio	Y	30	16	9	1,8	45,2	0,008	2,19722458	2010	Santiago	YY	5,5	3,37	8,81	0,0001	1,70474809	YY	1
hepatitis B vaccination	OR	MS	S	1	op	0	2001	Ascherio	N	51,7	6	0,9	0,5	1,6		0,10536052	2011	Farez	N	1	0,74	1,37	0,099	0	NN	1
KIF1B rs10492972 C/T allelic	OR	MS	G	0	ini	1	2008	Aulchenko	Y	35,2	1	1,34	1,2016	1,51	0,0001	0,29266961	2011	Kudryavtseva	NN	1,02	0,97	1,06	0,46	0,01980263	YY	0
IL7R T2441 rs6897932 CC/CT+TT	OR	MS	G	0	op	0	2007	Gregory	Y	35,2	29	1,29	1,14	1,46	0,0001	0,25464222	2011	Zhang	YY	1,15	1,06	1,24	0,0009	0,13976194	YY	1
IL7R T2441 rs6897932 C/T allelic	OR	MS	G	0	ini	1	2007	Heller	Y	51,7	32	1,18	1,09	1,27	0,0001	0,16553444	2011	Zhang	YY	1,11	1,04	1,19	0,001	0,10436002	YY	1
Smoking (prior to MS) conservative	RR	MS	S	1	op	0	2009	Hedstrom	Y	8,2	2	1,5	1,27	1,77	0,0001	0,40546511	2011	Handel	YY	1,48	1,35	1,63	0,0001	0,39204209	YY	1
hepatitis B vaccination	OR	MS	S	1	op	0	2004	Hernan	Y	8,2	10	3,1	1,5	6,3	0,002	1,13140211	2011	Farez	N	1	0,74	1,37	0,099	0	YN	0
Smoking (prior to MS) conservative	RR	MS	S	1	op	0	2005	Hernan	Y	9,9	1	1,3	1	1,69	0,049	0,26236426	2011	Handel	YY	1,48	1,35	1,63	0,0001	0,39204209	YY	1
Infectious mononucleosis	OR	MS	O	0	op	0	2005	Porsomby	Y	30	13	2,01	1,11	3,63	0,021	0,69813472	2010	Handel	YY	2,17	1,97	2,39	0,0001	0,77472717	YY	1
Smoking (prior to MS) conservative	RR	MS	S	1	op	0	2003	Rise	Y	8,2	19	1,81	1,1	2,9	0,014	0,39322685	2011	Handel	YY	1,48	1,35	1,63	0,0001	0,39204209	YY	1

Comments

- column B indicates how effect size were expressed: OR: odd ratio, d: standardized mean difference
- column D type of study: G: genetic, I: brain imaging, S: risk factor on which each subject can act, O: other
- column F ini: initial study, op: other primary study, meta: meta-analysis
- column I first author
- column J indicates whether the individual study reported a statistically significant effect. Y: yes, N: no
- column K IF: 2012 impact factor of the journal that published the individual study
- column L number of newspaper articles echoing each individual study
- column M ES: effect size
- columns N and O CI: confidence interval
- column P statistical significance (p value)
- column Q logarithm of the effect size
- column T first author of the meta-analysis corresponding to each individual study
- column U indicates whether the corresponding meta-analysis reported a statistically significant effect. Y: yes, N: no
- column V ES: effect size
- columns W and X CI: confidence interval
- column Y statistical significance (p value)
- column Z logarithm of the effect size
- column AA consistency: nominal agreement between each individual study and the corresponding meta-analysis (YNO: the individual study and the meta-analysis reported significant effects in the opposite direction)
- column AB replication validity: the individual study must be in nominal agreement with the corresponding meta-analysis and its effect size (ln log) must not be inflated by more than 100%
- marker indicates that this individual study has been published after another individual study on the same topic and which has been also echoed by newspapers
- columns I, J and L indicates that this individual study reported a null finding that has been echoed as such by newspapers
- columns L, J and L indicates that newspaper articles echoing this study put forward significant effects about other genetic factors, but did not report on this null finding

Parkinson																	ini 2									
Marker	# of lines		8			individual study				IF	media	ES	CI low	CI high	# pati	[LnES]	Meta-analysis									
	ES	path	type	S	order	ini	Year	Author	Sign							Year	Author	Sign	ES	CI low	CI high	p	[LnES]	consist	replic	
Pesticides exposure	RR	PD	S	1	op	0	2006	Ascherio	Y	11,2	16	1,8	1,3	2,5	413	1,323820192	2012	Van der Mark	YY	1,62	1,4	1,88	0,0001	1,03536904	YY	1
Head injury preceding c/c	RR	PD	O	0	op	0	2003	Bower	Y	8,2	10	4,3	1,2	5,5	196	1,458615	2012	Noyce	YY	1,58	1,3	1,91	0,0001	0,45742485	YY	0
Anti-infla non-asp-steroi all	RR	PD	S	1	ini	1	2003	Chen	Y	7,7	17	0,55	0,32	0,96	414	0,597837	2012	Noyce	Y	0,83	0,95	0,72	0,0084	0,18632958	YY	0
Smoking ever/never all	RR	PD	S	1	op	0	2010	Chen	Y	8,2	1	0,74	0,67	0,82	1662	0,3011051	2012	Noyce	YY	0,64	0,89	0,6	0,0001	0,4462871	YY	1
GBA gene N370S G/A cauca allelic	OR	PD	G	0	op	0	2007	Clark	Y	8,2	2	5,57	1,67	18,59	278	1,7173951	2012	Lill	YY	3,51	2,55	4,83	0,0001	1,25561604	YY	1
Statin use	RR	PD	O	0	op	0	2012	Gao	Y	7,7	1	0,74	1	0,54	644	0,3011051	2013	Undela	Y	0,77	0,92	0,64	0,0005	0,26136476	YY	1
LRRK2 G2019S rs34637584 allelic	OR	PD	G	0	op	0	2005	Gilks	Y	39	14	12,27			482	2,5071573	2012	Wu	YY	14,98	10,68	21,01	0,0001	2,70671598	YY	1
Well water exposure case/c	RR	PD	S	1	op	0	1990	Koller	Y	8,2	1	1,7	1,05	2,7	150	0,5306283	2012	Noyce	Y	1,21	1,04	1,4	0,012	0,19062036	YY	0
LRRK2 G2019S rs34637584 allelic	OR	PD	G	0	op	0	2006	Lesage	Y	51,7	10	29,31	6,96	123,5	104	3,3779285	2012	Wu	YY	14,98	10,68	21,01	0,0001	2,70671598	YY	1
LRRK2 G2019S rs34637584 allelic	OR	PD	G	0	op	0	2006	Ozelius	Y	51,7	11	17,6	5,9	52,2	120	2,8978989	2012	Wu	YY	14,98	10,68	21,01	0,0001	2,70671598	YY	1
Coffe drinking/non-drinking all	RR	PD	S	1	op	0	2000	Ross	Y	30	34	0,45	0,71	0,3	102	0,7885077	2012	Noyce	YY	0,67	0,76	0,58	0,0001	0,40047757	YY	1
LRRK2 R1628P rs33949390 allelic	OR	PD	G	0	ini	1	2008	Ross	Y	11,2	1	2,03	1,33	3,09	1230	0,7080358	2012	Wu	Y	1,96	1,13	3,41	0,02	0,67294447	YY	1
GBA gene N370S G/A cauca allelic	OR	PD	G	0	op	0	2009	Sidransky	Y	51,7	1	3,96	2,6	6,02	278	1,376244	2012	Lill	YY	3,51	2,55	4,83	0,0001	1,25561604	YY	1
Odor identification	d	PD	O	0	op	0	2001	Sobel	Y	9,7	1	2,06	1,29	2,83	20	3,736428	2012	Rayahel	YY	1,52	1,21	1,83	0,0001	2,756976	YY	1
Pesticides exposure	RR	PD	S	1	op	0	2009	Tanner	Y	7,7	4	1,9	1,12	3,21	519	0,6418539	2012	Van der Mark	YY	1,62	1,4	1,88	0,0001	0,48242615	YY	1
Smoking ever/never all	RR	PD	S	1	op	0	2007	Thacker	Y	8,2	1	0,71	0,58	0,87	413	0,3424903	2012	Noyce	YY	0,64	0,69	0,6	0,0001	0,4462871	YY	1
Anti-infla non-asp-steroi all	RR	PD	S	1	op	0	2007	Wahner	Y	8,2	15	0,52	0,35	0,79	293	0,6539265	2012	Noyce	Y	0,83	0,85	0,72	0,0084	0,18632958	YY	0

Comments

- column B indicates how effect size were expressed: OR: odd ratio, d: standardized mean difference
- column D type of study: G: genetic, I: brain imaging, S: risk factor on which each subject can act, O: other
- column F ini: initial study, op: other primary study, meta: meta-analysis
- column I first author
- column J indicates whether the individual study reported a statistically significant effect. Y: yes, N: no
- column K IF: 2012 impact factor of the journal that published the individual study
- column L number of newspaper articles echoing each individual study
- column M ES: effect size
- columns N and O CI: confidence interval
- column P statistical significance (p value)
- column Q logarithm of the effect size
- column T first author of the meta-analysis corresponding to each individual study
- column U indicates whether the corresponding meta-analysis reported a statistically significant effect. Y: yes, N: no
- column V ES: effect size
- columns W and X CI: confidence interval
- column Y statistical significance (p value)
- column Z logarithm of the effect size
- column AA consistency: nominal agreement between each individual study and the corresponding meta-analysis (YIO: the individual study and the meta-analysis reported significant effects in the opposite direction)
- column AB replication validity: the individual study must be in nominal agreement with the corresponding meta-analysis and its effect size (in log) must not be inflated by more than 100%
- marker indicates that this individual study has been published after another individual study on the same topic and which has been also echoed by newspapers

GLAUCOMA

initial 2

Marker	# of lines	2		0		individual study										Meta-analysis											
		ES	path	type	S	order	ini	Year	Author	Sign	IF	media	ES	CI low	CI high	p	[LnES]	Year	Author	Sign	ES	CI low	CI high	p	[LnES]	consist	replic
LOXL1 SNP rs2165241 exfol glauc. Caucas		OR	GLAU	G	0	ini	1	2007	Thorleifsson	Y	31	3	3.78	2.74	5.2	0,0001	1,32972401	2010	Chen	YY	3,39	3,07	3,74	0,0001	1,22082992	YY	1
optineurin Met98Lys AA+AT/TT POAG		OR	GLAU	G	0	ini	1	2002	Rezaie	Y	31	2	7.23	3,2697	15,983	0,0001	1,97823904	2010	Cheng	N	1,08	0,94	1,25		0,07696104	YN	0
meta-analysis echoed by newspapers																											
high myopia as risk of POAG		OR	GLAU	O		meta		2011	Marcus	Y	5,6	3	2,46	1,93	3,15	0,0001	0,90016135										

Comments

- column B indicates how effect size were expressed: OR: odd ratio, d: standardized mean difference
- column D type of study: G: genetic, I: brain imaging, S: risk factor on which each subject can act, O: other
- column F ini: initial study, op: other primary study, meta: meta-analysis
- column I first author
- column J indicates whether the individual study reported a statistically significant effect. Y: yes, N: no
- column K IF: 2012 impact factor of the journal that published the individual study
- column L number of newspaper articles echoing each individual study
- column M ES: effect size
- columns N and O CI: confidence interval
- column P statistical significance (p value)
- column Q logarithm of the effect size
- column T first author of the meta-analysis corresponding to each individual study
- column U indicates whether the corresponding meta-analysis reported a statistically significant effect. Y: yes, N: no
- column V ES: effect size
- columns W and X CI: confidence interval
- column Y statistical significance (p value)
- column Z logarithm of the effect size
- column AA consistency: nominal agreement between each individual study and the corresponding meta-analysis (YYO: the individual study and the meta-analysis reported significant effects in the opposite direction)
- column AB replication validity: the individual study must be in nominal agreement with the corresponding meta-analysis and its effect size (in log) must not be inflated by more than 100%
- marker indicates that this individual study has been published after another individual study on the same topic and which has been also echoed by newspapers

RHEUMATOID ARTHRITIS

Marker	# of lines	initial			largest			op			Meta			Meta-analysis			ES	CI low	CI high	p	[LnES]	consist	repro			
		ES	path	type	S	order	ini	Year	Author	Sign	IF	media	ES	CI low	CI high	p								[LnES]	Year	Author
PTPN22 C1858T rs2476601 T/C allelic	OR	RA	G	0	ini	1	2004	Begovich-2	Y	11.2	2	2.13	1.73	2.615	0.0001	0.756122	2012	Lee g	YY	1.64	1.514	1.77	0.0001	0.4947	YY	1
Smoking ever/never	OR	RA	S	1	op	0	2001	Hutchinson	Y	9.1	6	1.81	1.22	2.19	0.002	0.593327	2010	Sugiyama	YY	1.4	1.25	1.58	0.0001	0.3365	YY	1
FCRL3 -169 rs7528984 C/T allelic	OR	RA	G	0	ini	1	2005	Kochi	Y	35.2	1	1.37	1.1814	1.5965	0.0001	0.314811	2010	Lee d	N	1.06	0.987	1.146	0.107	0.0583	YN	0
TRAF1-C5 rs3761847 allelic	OR	RA	G	0	op	0	2009	Lee	N	7.6	1	0.98	0.85	1.08		0.040822	2010	Patsopoulos	YY	1.1	1.06	1.14	0.0001	0.0953	NY	0
Smoking ever/never	OR	RA	S	1	op	0	2001	Reckner-Olsson	Y	9.1	3	1.75	1.27	2.41	0.0006	0.559616	2010	Sugiyama	YY	1.4	1.25	1.58	0.0001	0.3365	YY	1
PTPN22 C1858T rs2476601 T/C allelic	OR	RA	G	0	op	0	2005	Simkins	Y	7.5	4	1.6	1.27	2.033	0.0001	0.470004	2012	Lee g	YY	1.64	1.514	1.77	0.0001	0.4947	YY	1
Smoking ever/never	OR	RA	S	1	op	0	2003	Stolt	Y	9.1	1	1.32	1.04	1.67	0.0216	0.277632	2010	Sugiyama	YY	1.4	1.25	1.58	0.0001	0.3365	YY	1
MHC2TA 168A/G rs3087456 G/A allelic A vs	OR	RA	G	0	ini	1	2005	Swanberg	Y	35.2	4	0.85	0.7541	0.9659	0.0121	0.162519	2008	Bronson	NN	0.95	0.901	1.03	0.062	0.0513	YN	0
TNFAIP3 rs6920220 A/G allelic	OR	RA	G	0	ini	1	2007	Thomson	Y	35.2	1	1.23	1.147	1.327	0.0001	0.207014	2012	Lee e	YY	1.22	1.16	1.26	0.0001	0.1989	YY	1
Smoking ever/never	OR	RA	S	1	op	0	2003	Tokuhiro	Y	35.2	14	1.25	1.07	1.46	0.049	0.223144	2008	Okada	YY	1.11	1.05	1.18	0.0008	0.1044	YY	0
SLC22A4 sisc2f1/F2 allelic	OR	RA	G	0	ini	1	2007	WTCCC	Y	38.6	32	1.12	1.029	1.213	0.0084	0.113328	2012	Song b	YY	1.08	1.029	1.131	0.002	0.077	YY	1

Comments

- column B indicates how effect size were expressed: OR: odd ratio, d: standardized mean difference
- column D type of study: G: genetic, I: brain imaging, S: risk factor on which each subject can act, O: other
- column F first author
- column J indicates whether the individual study reported a statistically significant effect. Y: yes, N: no
- column K IF: 2012 impact factor of the journal that published the individual study
- column L number of newspaper articles echoing each individual study
- column M ES: effect size
- columns N and O CI: confidence interval
- column P statistical significance (p value)
- column Q logarithm of the effect size
- column T first author of the meta-analysis corresponding to each individual study
- column U indicates whether the corresponding meta-analysis reported a statistically significant effect. Y: yes, N: no
- column V ES: effect size
- columns W and X CI: confidence interval
- column Y statistical significance (p value)
- column Z logarithm of the effect size
- column AA consistency: nominal agreement between each individual study and the corresponding meta-analysis (Y/O: the individual study and the meta-analysis reported significant effects in the opposite direction)
- column AB replication validity: the individual study must be in nominal agreement with the corresponding meta-analysis and its effect size (ln log) must not be inflated by more than 100%
- marker indicates that this individual study has been published after another individual study on the same topic and which has been also echoed by newspapers

Annexe 3

Annexe 4

Annexe 4

Enquête auprès des journalistes scientifiques

Liste des questions à poser lors d'interviews semi-directifs

1) Le journaliste et son champ d'activité

1.1) Êtes-vous journaliste attaché à un journal, une radio, une télévision ou bien êtes-vous pigiste indépendant? Dans le premier cas veuillez préciser pour quel média vous travaillez. Dans le deuxième, veuillez préciser les médias pour lesquels vous travaillez le plus souvent.

1.2) Quelle proportion de votre activité journalistique est dédiée à la production d'articles (ou documents audiovisuels) faisant état de résultats de la recherche scientifique ou biomédicale?

1.3) Si vous le voulez bien pourriez-vous indiquer votre âge (ou votre tranche d'âge)? Depuis combien d'années travaillez-vous en tant que journaliste scientifique?

1.4) Au cours de votre formation universitaire, avez-vous reçu un enseignement scientifique? Si oui, dans quel domaine et de quel niveau?

1.5) Quand vos articles abordent des questions scientifiques, traitent-ils des domaines suivants (précisez les proportions) ?

- biomédical, biologie humaine, pharmacie
- physique
- chimie
- biologie animale et végétale
- sciences de la terre (climat, etc)
- sciences de l'ingénieur (informatique, etc)
- sciences humaines

1.6) En dehors de votre activité de journaliste indépendant, êtes-vous aussi rémunéré pour des activités de communication? Si oui, pouvez-vous nous indiquer pour quelles compagnies?

1.7) Pensez-vous qu'il peut y avoir conflit d'intérêt entre l'activité de communicant et celle de journaliste? Pour vous-même? Pour d'autres journalistes de votre connaissance?

2) Le choix du sujet traité

2.1) Qu'est-ce qui détermine le choix du sujet que vous traitez dans un article? Pourriez-vous préciser dans quelle proportion l'impulsion vient:

- de vous-même,
- de la direction du journal qui vous a commandé l'article
- d'un scientifique (médecin y compris) qui a pris l'initiative de vous contacter
- d'une institution scientifique qui vous a contacté
- d'un événement social ou politique (par exemple la publication d'une recommandation de l'HAS)

2.2) Lorsque l'impulsion vient de vous-même, est-ce suite à vos lectures, et si oui lesquelles?

- des journaux scientifiques primaires
- des communiqués de presse rédigés par les grandes revues scientifiques

- des communiqués de presse rédigés par les services de presse des institutions scientifiques (CNRS, universités, etc)
- des articles rédigés par les agences de presse (e.g. AFP)
- des autres articles de la presse grand public

3) Préparation d'un document médiatique contenant des informations scientifiques

3.1) Pour préparer votre article ou document audiovisuel, sur quelles sources écrites vous appuyez-vous? En particulier vérifiez-vous vous-même l'information scientifique en vous référant aux journaux scientifiques primaires?

3.2) Recherchez-vous l'avis des scientifiques qui ont produit l'information (i.e. les auteurs des articles primaires), l'avis d'autres scientifiques experts du domaine?

3.3) Lorsque vous citez des scientifiques, leur demandez-vous de valider leurs propos avant publication?

3.4) Lorsque vous citez une étude scientifique précisant que les auteurs ont un conflit d'intérêt de nature financière, le mentionnez-vous dans votre article?

3.5) Vous arrive-t-il de renoncer à rédiger un article? Si oui, pour quelles raisons?

3.6) De combien de temps disposez vous pour préparer un article ou un document audiovisuel?

3.7) Le document médiatique que vous avez préparé est-il parfois (ou souvent?) coupé ou remanié par la rédaction?

3.8) Qui décide du titre de l'article dans sa version finale: vous, la rédaction, les deux en concertation?

4) Position du journaliste par rapport à la recherche

4.1) A votre avis, dans le domaine de la recherche biomédicale, quelle est la proportion de découvertes initiales qui ont été soit réfutées soit fortement atténuées par les études ultérieures sur la même question?

4.2) Estimez-vous important d'informer le public lorsqu'une information scientifique qui avait été médiatisée a ensuite été réfutée par les études ultérieures? Si oui, avez vous déjà rédigé un ou des articles mentionnant une réfutation? Si oui, combien?

4.3) A votre avis, lorsque des scientifiques s'adressent à vous, tiennent-ils le même discours que lorsqu'ils s'adressent à leurs pairs?

5) Avez-vous un commentaire à ajouter?

Survey for scientific journalists

Semi-directive interviews

1) The journalist and its scope

1.1) Are you working as a freelance journalist or for a newspaper, a radio or a TV channel? For each of these cases, please give me some details as to the type of media you are working for the most often.

1.2) How much of your work is dedicated to the preparation of articles or broadcast reporting on scientific or biomedical research?

1.3) Would you mind telling us your age (or your age range)? How long have you been working as a journalist?

1.4) Do you have a degree in science? If you do, in which area? What is your level of qualification?

1.5) Which domains of science are you reporting on? In which proportion?

- biomedical science, human biology, pharmacy
- physics
- chemistry
- Earth science (climate)
- engineering (computing)
- human sciences

1.6) Do you happen to work sometimes as a communication consultant? If you do, which companies are you working for?

1.7) Do you think there might be a conflict of interest for someone working as a journalist and a communication consultant as well? For yourself? For some colleagues you might know?

2) Story selection

2.1) How do you pick your stories? Where does the idea come from?

- yourself
- the editor of the newspaper you are working for
- a scientist or a MD contacting you
- a science institution,
- a company contacting you
- a social or political event (such as a WHO report)

2.2) When you are the one picking your story, does it occur after you read something about it? If so, where?

- in scientific journals
- press releases from scientific journals
- press releases from institutions (e. g. universities)
- press releases from news agencies (e.g. Associated Press)
- company press releases

- stories in newspapers

3) Science reporting

3.1) What sources do you use to report on science? Do you examine the original scientific observations in the scientific journals themselves?

3.2) Do you interview the authors of the scientific publication or some experts in the field?

3.3) When you cite scientists do you ask them to check your story before its publication?

3.4) If you are reporting on a scientific study and its authors have a financial conflict of interest, do you include this information in your paper?

3.5) Have you ever decided not to write an article? If so, why?

3.6) How long do you have to write a paper or prepare a TV or radio broadcast?

3.7) Is your article or broadcast shortened or reworked by the editor? If so how often?

3.8) Who picks the headline? You, your editor or both of you?

4) Journalist's opinion about research.

4.1) In your opinion, what is the rate of initial studies that have been refuted or strongly attenuated by subsequent studies in biomedical research?

4.2) When a scientific study has been covered by the media and is later refuted, do you think laypeople should know about it? If so, have you ever written such a refutation story? How many times?

4.3) Do you think that scientists use the same discourse when talking to their peers or to you?

5) Is there anything else you wish to add?

Annexe 5

Annexe 5

Donnez votre opinion :

Mesures proposées pour améliorer la reproductibilité des recherches biomédicales.

Nous vous demandons de répondre à cette enquête préliminaire afin de nous aider à préparer une enquête plus large dans le cadre d'une thèse de doctorat en sciences politiques et concernant la reproductibilité des recherches biomédicales. Merci d'avance.

Estelle Mallet et Andy Smith (Centre Emile Durkheim, Université de Bordeaux)

Merci de coder votre opinion en cochant la colonne choisie et si vous n'avez pas d'opinion sur une mesure, ne cochez rien.

+ : la mesure vous semble efficace

- : la mesure ne vous paraît pas efficace

± : la mesure serait peut être efficace, mais elle vous paraît inapplicable

10 mesures pour améliorer la reproductibilité	+	-	±
Le protocole expérimental, les hypothèses et les méthodes d'analyse des données devraient être déposées sur un site officiel et publiquement accessible, avant de commencer à collecter les données.			
Les résultats d'une expérience devraient être observés à l'aveugle .			
Les données brutes devraient être accessibles sur un site officiel en libre accès.			
Un résultat devrait être considéré comme significatif seulement à $p < 0,005$			
Arrêter de financer la recherche sur projet et financer les équipes à postériori			
Ne pas tenir compte du facteur d'impact quand on juge une équipe ou un chercheur			
Encourager les études de réplication			
Encourager les recherches collaboratives et multicentriques			
Encourager la publication des observations négatives (contredisant l'hypothèse)			
Pour chaque article, rendre public le nom des reviewers			

Veillez rendre votre questionnaire à l'accueil ou aux organisateurs des ateliers. Vous pouvez aussi le scanner et l'envoyer à l'adresse ci-dessous. Si vous souhaitez ajouter un commentaire, n'hésitez pas et utilisez le reste de cette feuille.

estelle.mallet@u-bordeaux.fr