



HAL
open science

Méthodologie et qualité du reporting des études en oncologie (essais de phase II et études observationnelles)

Romain Rivoirard

► To cite this version:

Romain Rivoirard. Méthodologie et qualité du reporting des études en oncologie (essais de phase II et études observationnelles). Médecine humaine et pathologie. Université de Lyon, 2019. Français. NNT : 2019LYSE1151 . tel-02724796v2

HAL Id: tel-02724796

<https://theses.hal.science/tel-02724796v2>

Submitted on 2 Jun 2020

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.

N° d'ordre NNT : xxx



THESE de DOCTORAT DE L'UNIVERSITE DE LYON

Opérée au sein de
L'Université Claude Bernard Lyon 1

Ecole Doctorale N° 205

Ecole Doctorale Interdisciplinaire Sciences-Santé (EDISS)

Spécialité de doctorat : Médecine

Discipline : Méthodologie

Soutenue publiquement le 13/09/2019, par :

Romain RIVOIRARD

Méthodologie et qualité du reporting des études en oncologie (essais de phase II et études observationnelles)

Devant le jury composé de :

Madame BOURMAUD Aurélie, Maître de Conférence Universitaire-Praticien Hospitalier (MCU-PH), Université Paris Diderot, **Co-directrice de thèse**

Monsieur CHAUVIN Franck, Professeur des Universités-Praticien Hospitalier (PU-PH), Université Jean Monnet Saint-Etienne - Université Claude Bernard Lyon1, **Directeur de thèse**

Madame FALANDRY Claire, Professeur des Universités-Praticien Hospitalier (PU-PH), Université Claude Bernard Lyon1, **Examinatrice**

Madame LAPORTE Sylvie, Maître de Conférence Universitaire-Praticien Hospitalier (MCU-PH), Université Jean Monnet Saint-Etienne, **Rapporteuse**

Monsieur MASSARD Christophe, Praticien Hospitalier (PH), Institut Gustave Roussy, **Rapporteur**

Monsieur MERROUCHE Yacine, Professeur des Universités-Praticien Hospitalier (PU-PH), Université de Reims, **Examineur**

Résumé de thèse en français

Le cancer demeure en France un problème majeur de santé publique : Cette pathologie reste la principale cause de mortalité, avec près de 150 000 décès chaque année. Afin de lutter contre cette maladie, le plan cancer 2014-2019 a posé comme objectif de développer la recherche clinique, en incluant dès 2019 au moins 50 000 patients par an dans des essais thérapeutiques.

La recherche clinique en oncologie est en plein essor depuis l'avènement de thérapies telles que les thérapies moléculaires ciblées ou l'immunothérapie. Le nombre de nouveaux médicaments en développement en oncologie a augmenté de façon exponentielle depuis les années 2000. Il en découle un accroissement important du nombre de publications d'études en oncologie. L'analyse d'une étude n'est possible que si son schéma expérimental, sa réalisation et ses conditions d'analyse sont clairement décrits dans l'article publié. Le reporting est le fait de décrire dans une publication toutes les actions, notamment d'ordre méthodologiques, réalisées au cours de l'étude.

4 revues systématiques de la littérature ont été réalisées afin d'analyser le reporting des articles d'études en oncologie. La première étude a objectivé que la littérature était pauvre en articles analysant le reporting des études observationnelles (2% des publications) et que le reporting intrinsèque des revues systématiques récentes portant sur le sujet était exhaustif (score OQSR moyen (2010-2014) = 10, écart-type = 2). La deuxième revue systématique a mis en évidence un manque de données sur le reporting des essais de phase II en oncologie (13 articles) et 3 scores d'évaluation du reporting avaient été décrits : score OQS, score KMS et score d'index de qualité. 826 articles d'études ont été analysés dans la troisième revue systématique, dont 84.5% d'études observationnelles. Les reporting des variables et du critère de jugement principal étaient clairs, respectivement dans 100% et 85.2% des cas. Une discordance de reporting pour le test statistique principal utilisé, entre les parties Méthodes et Résultats, a été identifiée dans 23 articles. 557 publications d'essais de phase II ont été étudiées dans la quatrième revue systématique. Le score de reporting KMS était égal à 0, 1, 2, 3 respectivement pour 3,9%, 21,4%, 48,5% et 26,2% articles. En analyse multivariée, les 3 variables associées à la valeur du score KMS étaient : Le reporting clair du design statistique (OR = 2.22; IC 95% [1.36-3.65] ; p < 0.001), l'essai monocentrique (OR = 0.25 [0.09 – 0.74] ; p = 0.012) et l'analyse menée per protocole (OR = 0.48; IC 95% [0.32 -0.72] ; p < 0.001).

En conclusion, le reporting des études en oncologie reste perfectible. Les points d'optimisation relevés sont l'élaboration de recommandations de reporting spécifique aux essais de phase II en oncologie et l'adhésion de l'ensemble des journaux médicaux aux différentes recommandations de reporting.

Mots clés : Oncologie ; méthodologie ; reporting ; revue de littérature ; essais de phase II ; études observationnelles

Intitulé et adresse de l'unité ou du laboratoire : Université de Lyon, Université Lyon 1 - Laboratoire Health Services and Performance Research- HESPER-EA 7425, Site Lacassagne 162 Avenue Lacassagne 69424 LYON CEDEX 03

Résumé de thèse en anglais

Cancer remains a major public health problem in France: This pathology is the leading cause of death, with nearly 150,000 deaths each year. In order to fight against this disease, the 2014-2019 French cancer plan set the goal of developing clinical research, including at least 50,000 patients per year in 2019 in therapeutic trials.

Clinical research in oncology has been booming since the advent of therapies such as targeted molecular therapies or immunotherapy. The number of new drugs being developed in oncology has increased exponentially since the 2000s, resulting in a significant increase in the number of oncology study publications. The analysis of a study is possible only if its experimental design, its realization and its conditions of analysis are clearly described in the published article. Reporting is the process of describing in a publication all the actions, including methodological ones, carried out during the study.

4 systematic reviews were conducted to analyze the reporting of oncology study articles. The first study showed that the literature was poor in articles analyzing the reporting of observational studies (2% of publications) and that the intrinsic reporting of recent systematic reviews on the subject was exhaustive (average OQSR score (2010-2014) = 10, standard deviation = 2). The second systematic review revealed a lack of data on the reporting of phase II trials in oncology (13 articles) and 3 reporting evaluation scores were described: OQS, KMS and quality index score. 826 articles of study were analyzed in the third systematic review, 84.5% of which were observational studies. The reporting of the variables and the primary endpoint were clear in 100% and 85.2% of the cases, respectively. A reporting discrepancy for the main statistical test used, between the Methods and Results parts, was identified in 23 articles. 557 phase II trial publications were reviewed in the fourth systematic review. The KMS reporting score was 0, 1, 2, 3 respectively for 3.9%, 21.4%, 48.5% and 26.2% publications. In multivariate analysis, the 3 variables associated with the KMS score were : Reporting of a clear statistical design (OR = 2.22, 95% CI [1.36-3.65], $p < 0.001$), the single-center trial (OR = 0.25 [0.09 - 0.74], $p = 0.012$) and per protocol analysis (OR = 0.48, 95% CI [0.32-0.72], $p < 0.001$).

In conclusion, reporting of studies in oncology remains perfectible. Optimization points identified are the development of specific reporting recommendations for phase II trials in oncology and the adherence of all medical journals to the various reporting recommendations.

Keywords: Oncology; methodology; reporting; literature review ; phase II trials; observational studies

Table des matières

I- Introduction	page 6
1- Epidémiologie du cancer dans le monde, en Europe, en France	page 6
2- Les essais cliniques en oncologie et leurs spécificités méthodologiques	page 7
2-1 Essais de phase I	page 7
2-1-1 Objectifs et population étudiée	page 7
2-1-2 Plan de traitement	page 8
2-2 Essais de phase II	page 16
2-2-1 Objectifs	page 16
2-2-2 Méthodologie	page 16
2-3 Essais de phase III	page 22
2-3-1 Objectifs	page 22
2-3-2 Randomisation	page 22
2-3-3 Critère principal	page 23
2-3-4 Types d'essais	page 23
2-3-5 Plan d'expérience	page 25
2-3-6 Conclusion	page 25
3- Le reporting des études et son évaluation	page 26
3-1 Histoire du reporting	page 26
3-1-1 Un reporting incomplet et peu précis des essais cliniques	page 26
3-1-2 La déclaration CONSORT	page 27
3-1-3 La déclaration STROBE	page 28
3-1-4 La déclaration PRISMA	page 29
3-1-5 Les autres déclarations	page 30
3-2 Evaluation de la qualité de reporting des études	page 31
3-2-1 Evaluation globale de la qualité du reporting des études	page 32
3-2-2 Evaluation de la qualité du reporting des études en dentisterie	page 35
3-2-3 Evaluation de la qualité du reporting des études en acupuncture	page 36
3-2-4 Evaluation de la qualité du reporting des études en pédiatrie	page 37
3-2-5 Evaluation de la qualité du reporting des études en oncologie	page 38
II- Hypothèses de recherche et objectifs de la thèse	page 40
III- Les articles originaux évaluant le reporting en oncologie	page 42
1- Quality of reporting in oncology studies : A systematic analysis of literature reviews and prospects	page 42

Table des matières (suite)

1-1 Introduction	page 42
1-2 Méthodes	page 42
1-3 Résultats	page 43
1-4 Discussion	page 44
1-5 Article original	page 45
2- Evaluation of the quality of the reporting of phase II clinical trials in oncology: A systematic review	page 56
3-1 Introduction	page 56
3-2 Méthodes	page 56
3-3 Résultats	page 57
3-4 Discussion	page 58
3-5 Article original	page 60
3- Outcomes definitions and statistical tests in oncology studies: A systematic review of the reporting consistency	page 66
2-1 Introduction	page 66
2-2 Méthodes	page 66
2-3 Résultats	page 68
2-4 Discussion	page 69
2-5 Article original	page 71
4- Quality of reporting in oncology phase II trials : A 5-Year assesment through systematic review	page 84
4-1 Introduction	page 84
4-2 Méthodes	page 84
4-3 Résultats	page 86
4-4 Discussion	page 87
4-5 Article original	page 90
IV- Discussion	page103
4-1 Un plaidoyer pour des recommandations internationales	page106
4-2 Quel lien entre qualité du reporting et efficacité d'un médicament ?	page115
V- Conclusion	page120
VI- Bibliographie	page121

I- Introduction

1- Epidémiologie du cancer dans le monde, en Europe, en France

Le cancer constitue depuis plusieurs années un problème de santé majeur à l'échelle mondiale : En 2015, il s'agissait de la deuxième cause de décès dans le monde, avec 8.8 millions de morts en 2015 [1]. Il a été estimé qu'un décès sur 6 dans le monde est dû à une pathologie cancéreuse [1]. Il existe des facteurs de risque comportementaux et alimentaires bien connus (un indice élevé de masse corporelle, une faible consommation de fruits et légumes, le manque d'exercice physique, le tabagisme et la consommation d'alcool principalement [2]), mais, malgré ces connaissances, seul un pays sur cinq à revenu faible ou intermédiaire dispose des données nécessaires pour conduire une politique de lutte efficace contre le cancer [3]. L'accès à un diagnostic précis du cancer et à un traitement spécifique reste encore très inégal entre les pays à revenu faible et les pays à revenu élevé : moins de 30% des pays à revenu faible déclarent disposer de service de traitement, contre plus de 90% des pays à revenu élevé [1]. Tous ces éléments illustrent, à l'échelle mondiale, l'accès inégal de l'Homme aux thérapies spécifiques du cancer et à l'innovation thérapeutique. L'impact économique du cancer représente aussi un problème majeur à l'échelle mondiale : Ainsi, en 2010, il a été estimé à quelque 1160 milliards de dollars (US \$) le coût annuel total de la maladie [4].

A l'échelle Européenne, en termes d'incidence, l'Age Standardised Rate ou ASR en 2012, était estimé à 735.8 nouveaux cas pour 100 000 habitants chez l'homme et à 479.01 nouveau cas pour 100 000 habitants chez la femme [5]. Concernant la mortalité par cancer en 2012, l'ASR était estimé à 381.9 décès/100 000 habitants chez l'homme et à 212.03 décès/100 000 habitants chez la femme [5].

Au niveau de la France métropolitaine, on estime à 399 500 le nombre de nouveaux cas de cancer diagnostiqués en 2017 (214 000 hommes et 185 500 femmes) [6]. En 2015, l'âge médian au diagnostic était assez comparable entre les hommes et les femmes : 68 ans chez l'homme et 67 ans chez la femme. En termes d'incidence, les cancers de prostate, du poumon et colorectal sont les plus fréquents chez l'homme alors que chez la femme on retrouve le cancer du sein, colorectal et broncho-pulmonaire [6]. En France, les tumeurs malignes représentent la principale cause de décès chez l'homme, et la deuxième cause chez la femme, après les maladies de l'appareil circulatoire, en 2013 [7].

Le cancer demeure donc une pathologie grave, fréquente, avec un fardeau croissant. Il est donc important d'améliorer son pronostic, et notamment en développant de nouvelles thérapeutiques, tout en contraignant autant que faire se peut les coûts. En France, le plan cancer 2014-2019 a, entre autre, posé comme objectif de développer la recherche clinique, en incluant dès 2019 au moins 50 000 patients par an dans des essais thérapeutiques conduits par des industriels ou par des académiques (hôpitaux publics, privés, organisation de recherche, INSERM...) [8]. Cela correspond à un doublement du nombre d'inclusions (25 000) par rapport à 2013.

2- Les essais cliniques en oncologie et leurs spécificités méthodologiques

Depuis 1989, l'oncologie médicale est reconnue en France comme spécialité d'exercice [9]. Les médecins spécialistes en oncologie médicale sont en première ligne dans la prise en charge du patient atteint de cancer, ils interviennent à la croisée des différents médecins spécialistes (gynécologues, urologues, pneumologues, gastro-entérologues, chirurgiens thoraciques et digestifs ...) dans la stratégie thérapeutique globale de chaque patient atteint de cancer. Ils expliquent et détaillent le programme personnalisé de soins aux patients et restent très impliqués en recherche clinique.

Avec l'avènement récent de thérapies innovantes, telles que les Thérapies Moléculaires Ciblées (TMC) ou l'immunothérapie, le nombre de nouveaux médicaments en développement en oncologie a augmenté de façon exponentielle depuis les années 2000 [10].

En oncologie, la recherche clinique est centrée sur l'évaluation de nouvelles modalités de prise en charge thérapeutique de personnes atteintes de cancer, mais elle concerne également la prévention et les méthodes diagnostiques. L'essai clinique constitue une étape incontournable afin de déterminer l'efficacité de ces interventions. Nous allons nous concentrer dans ce travail de thèse sur les essais cliniques portant spécifiquement sur l'évaluation des médicaments

Au cours du développement d'un médicament, différentes étapes d'apprentissage et de confirmation d'informations vont se succéder [11], qui seront par la suite résumées et soumises aux agences d'autorisation de mise sur le marché. L'essai clinique d'oncologie comprend quatre phases distinctes (phase I, phase II, phase III avec éventuelle obtention au décours d'une Autorisation de Mise sur le Marché ou AMM et phase IV dite phase « post AMM »), toujours précédées d'une phase pré clinique. Chacune de ces phases a un objectif et des particularités méthodologiques distincts.

2-1 Essais de phase I

2-1-1 Objectifs et population étudiée

La phase I correspond à la première utilisation clinique d'un nouveau médicament sur l'Homme. En oncologie, les médicaments analysés en phase I ont été préalablement sélectionnés du fait de leur activité sur des tumeurs d'origine animale ou humaine lors de la phase pré clinique. La spécificité des essais de phase I en oncologie est que, le plus souvent, le médicament étudié (agent cytotoxique ou TMC à fort potentiel toxique) est administré à des patients malades, porteurs de tout type de tumeurs solides ou hématologiques, ayant échappé aux thérapeutiques conventionnelles disponibles, ou à des patients atteints de tumeurs pour lesquelles il n'existe aucun traitement validé. Les patients éligibles dans ces essais doivent être le plus souvent en bon état général (Performans Status 0 ou 1, indice Karnofsky >70%), sans défaillance viscérale et avec une espérance de vie estimée supérieure à trois mois. Sur le plan éthique, ces essais sont planifiés de façon à minimiser le nombre de patients traités à des doses non thérapeutiques. Comme dans les autres spécialités médicales, l'essai de phase I en oncologie s'intéresse au profil pharmacocinétique du médicament après administration

(absorption, distribution, métabolisme et élimination), au profil de toxicité et de sécurité d'emploi mais aussi à établir la dose optimale à administrer pour l'essai de phase II ultérieur.

L'approche utilisée diffère selon le type de médicament à l'étude : Pour les agents cytotoxiques, la courbe dose-réponse (efficacité/toxicité) est le plus souvent croissante monotone. Du fait de leur mécanisme d'action non spécifique, les données de toxicités sont primordiales pour ces médicaments, afin d'établir les courbes dose-toxicité. La définition de la Toxicité Dose-Limitante (TDL) est ainsi la première étape nécessaire pour obtenir ces données de toxicité. La TDL peut être composée d'un seul élément toxique ou une combinaison d'événements (exemples : « toute toxicité hématologique irréversible de grade supérieur ou égal à deux » ou « nausées de grade trois ou neutropénie de grade trois persistant pendant plus de deux semaines »). Une fois la TDL déterminée, l'objectif principal consistera à estimer la Dose Maximale Tolérée (DMT), qui est la dose du médicament correspondant au risque prédéfini de toxicité sévère. Il s'agit donc d'un protocole d'escalade de dose, pour lequel il sera nécessaire de fixer précisément la dose de départ ainsi que le mode d'escalade de dose.

Dans le cas particulier des TMC, ces médicaments sont supposés être moins toxiques que les agents cytotoxiques, du fait de leur mécanisme d'action ciblant un mécanisme précis de l'oncogénèse. Ainsi, pour ces nouvelles thérapeutiques, les études de phase I peuvent dans certains cas être réalisées, non pas chez des patients atteints de cancer, mais chez des sujets sains, lorsque les données précliniques chez l'animal ont montré un profil de toxicité favorable. L'erlotinib, inhibiteur de tyrosine kinase utilisé dans l'adénocarcinome pulmonaire muté EGFR, est un exemple concret de ce type de développement [12].

2-1-2 Plan de traitement

Il convient de préciser que la méthodologie détaillée par la suite concerne les médicaments cytotoxiques et cytostatiques présentant un risque majeur de toxicité, ce qui représente une grande partie des thérapeutiques testées en oncologie. Deux données sont indispensables pour la planification de l'essai de phase I : La dose de départ et la méthode d'escalade de dose.

a/ Dose de départ

A ce stade de l'étude, il n'est pas éthique d'encourir le moindre risque de toxicité chez l'Homme. Ainsi, la dose de départ est le plus souvent fixée comme étant le dixième de la DL_{10} si elle n'est pas toxique chez le chien. La DL_{10} ou Dose Létale 10% correspond à la dose provoquant la mort de 10% de la population de souris exposée à la substance en étude préclinique. En cas de toxicité induite chez le chien, la dose de départ chez l'Homme correspond à un tiers de la plus faible dose induisant une toxicité chez l'animal.

b/ Méthodes d'escalade de dose

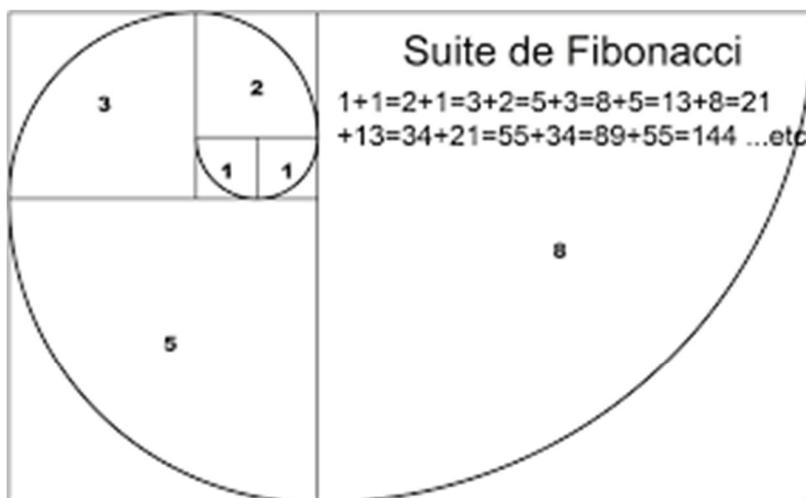
En oncologie, les essais de phase I diffèrent essentiellement sur la méthode d'escalade de dose, le nombre de patients par palier de dose, l'affectation d'un patient à un palier de dose

ainsi que le critère d'arrêt d'escalade de dose. Les différents modes d'escalade de dose sont détaillés ci-dessous.

Méthode selon la suite de Fibonacci modifiée

Cette méthode d'escalade de dose s'inspire de la séquence historique de Fibonacci, mathématicien italien du 13^{ème} siècle, où chaque nombre est la somme des deux précédents : 0 et 1 > 1+0=1, puis 1 et 1 > 1+1=2, puis 1 et 2 > 1+2=3, puis 2 et 3 > 2+3=5, puis 3 et 5 > 3+5=8, (**Figure 1**) etc.

Figure 1 : La suite de Fibonacci



Il est important de noter que les bases scientifiques justifiant l'utilisation de cette méthode d'escalade de dose en oncologie restent discutables. La problématique de cette méthode et que, lors de l'escalade de dose, les paliers (ou « marches ») successifs sont de plus en plus importants et le patient peut donc être exposé à des toxicités importantes à des paliers élevés. Afin de limiter ce problème, cette suite de Fibonacci a été modifiée, dans le but de réduire la hauteur des paliers de doses les plus élevés. **Le tableau 1** illustre les paliers de dose selon la suite de Fibonacci et selon la suite de Fibonacci modifiée. Il faut toutefois être vigilant, car la modification de la suite peut varier selon les essais. Par soucis de transparence et de reproductibilité, il semble donc préférable de rapporter précisément la suite utilisée, plutôt que de noter seulement « selon le schéma de Fibonacci modifié ».

Dans le schéma standard, des cohortes de 3 patients par palier de dose sont inclus. Les doses sont le plus souvent exprimées en mg/m^2 . L'escalade de dose est uniquement réalisée interpatients. Le processus se poursuit jusqu'à la survenue d'un pourcentage prédéfini d'une Toxicité Dose-Limitante (TDL).

Les règles de décision sur la poursuite ou l'arrêt de l'escalade de dose sont le plus souvent prédéfinies ainsi (**Figure 2**) :

- Si aucune TDL n'est observée chez les 3 patients de la cohorte, l'escalade continue.

- Si une TDL est observée chez un seul patient de la cohorte, 3 patients supplémentaires devront être inclus au même palier. Si aucun de ces nouveaux patients ne présente une TDL, on continue l'escalade, sinon l'escalade est arrêtée.
- Si une TDL est observée chez 2 ou 3 patients de la cohorte, l'escalade de dose est arrêtée.

Une fois l'escalade arrêtée, 3 derniers patients sont inclus au palier inférieur précédant le palier d'arrêt. La Dose Maximale Tolérée (DMT) est ainsi définie comme le palier précédant la dose d'arrêt de l'escalade et permet de définir la dose recommandée pour les essais de phase II ultérieurs.

Tableau 1 : Escalade de dose selon la suite de Fibonacci et selon la suite modifiée (dose de départ 1mg/m²)

Palier	Série de Fibonacci		Série de Fibonacci modifiée	
	Dose (mg/m ²)	Incrément (%)	Dose (mg/m ²)	Incrément (%)
1	1	-	1	-
2	2	100	2	100
3	3	50	3,3	67
4	5	67	5	50
5	8	60	7	40
6	13	63	9	33
7	21	62	12	33
8	34	62	16	33

Ce schéma traditionnel, encore largement utilisé en oncologie, présente plusieurs inconvénients:

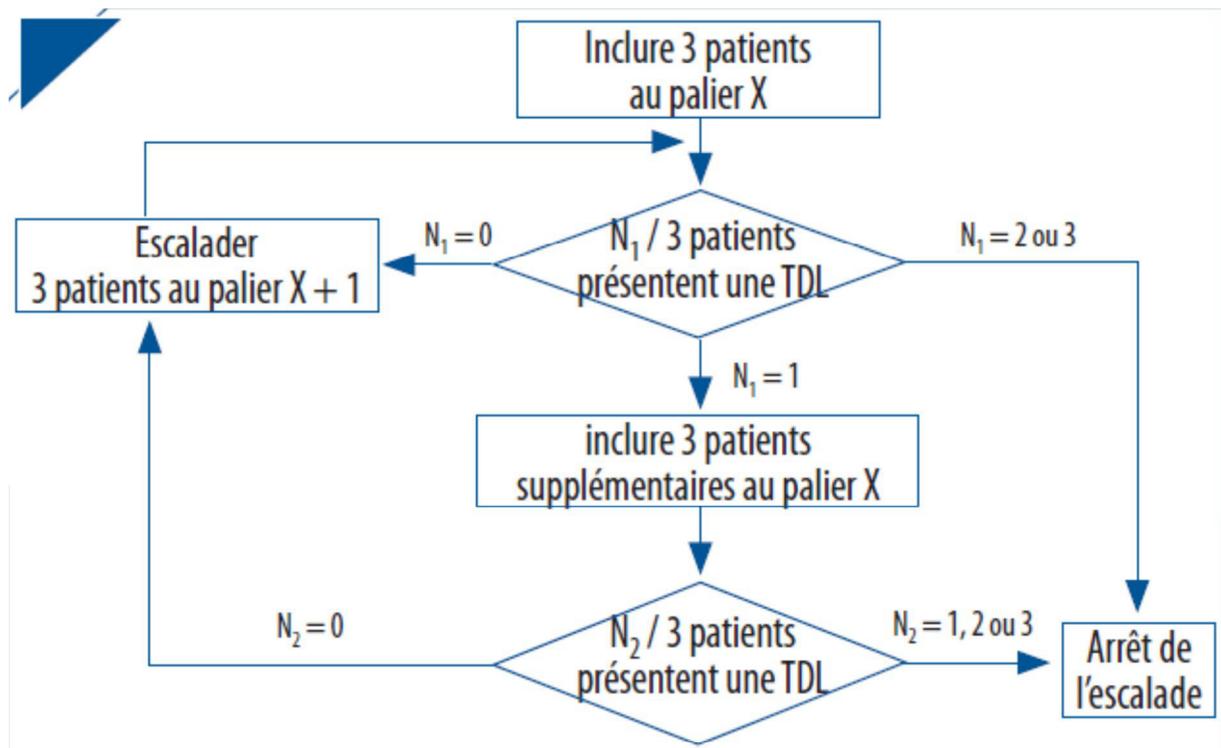
Le premier est d'ordre éthique : Avec ce schéma d'escalade, trop de patients reçoivent des doses infra thérapeutiques quand la dose de départ est loin de la DMT. De plus, en prenant en compte que l'escalade de dose intrapatient n'est pas autorisée avec ce schéma, les patients commençant par exemple au palier 1 infra thérapeutique recevront toujours une dose non thérapeutique.

Le deuxième est que chaque étape ne tient aucunement compte de l'information recueillie au palier précédent.

Le troisième est que, selon les études réalisées, il est possible de ne jamais atteindre la DMT.

Enfin, le quatrième et l'un des plus importants, est que la DMT obtenue avec ce schéma manque cruellement de précision. En effet, la DMT obtenue avec de tels essais représente la dose pour laquelle le pourcentage de survenue d'une TDL est compris entre 15 et 70%. Ces pourcentages sont calculés sur seulement 3 patients d'une cohorte et ceci n'est pas acceptable sur le plan statistique.

Figure 2 : Schéma standard d'escalade de dose (schéma de Fibonacci modifié), affectation des patients et critères d'arrêt



Méthode d'escalade accélérée

La méthode d'escalade accélérée (ou Accelerated Titration Design, ATD) est une méthode qui a été conçue afin de contourner certaines limites de la méthode selon un schéma de Fibonacci [13]. Avec cette méthode, une première étape consiste à réaliser une escalade de dose accélérée : Un seul patient est inclus par palier de dose jusqu'à la survenue d'une TDL et les doses sont doublées de palier en palier. Dès l'apparition d'une TDL ou d'une toxicité quelconque de grade 2, 6 ou 3 patients sont respectivement inclus à ce palier. La deuxième étape consiste à continuer l'escalade de dose selon le schéma classique décrit plus haut. Il est important de préciser qu'avec cette méthode, l'escalade de dose intra patient est possible.

Cette approche permet ainsi de réduire de façon significative le nombre de patients traités à des doses infra thérapeutiques du fait, d'une part, du nombre réduit de patients aux faibles paliers de dose pour chaque cohorte et, d'autre part, de l'escalade de dose chez un même patient, augmentant sa chance d'obtenir une dose efficace. Cette méthode permet aussi d'obtenir des données de toxicité cumulatives. Cependant, cette escalade de dose est plus

« agressive », avec un risque plus important de toxicité pour le patient. Il est donc essentiel de bien définir au préalable ce qui est acceptable comme risque de toxicité pour le patient. Malgré tout, cette méthode ne permet pas de résoudre le problème de la détermination de la DMT avec une bonne précision. D'autres méthodes alternatives ont été développées, parmi elles, la plus connue est celle de la réévaluation continue.

Méthode par réévaluation continue (ou CRM)

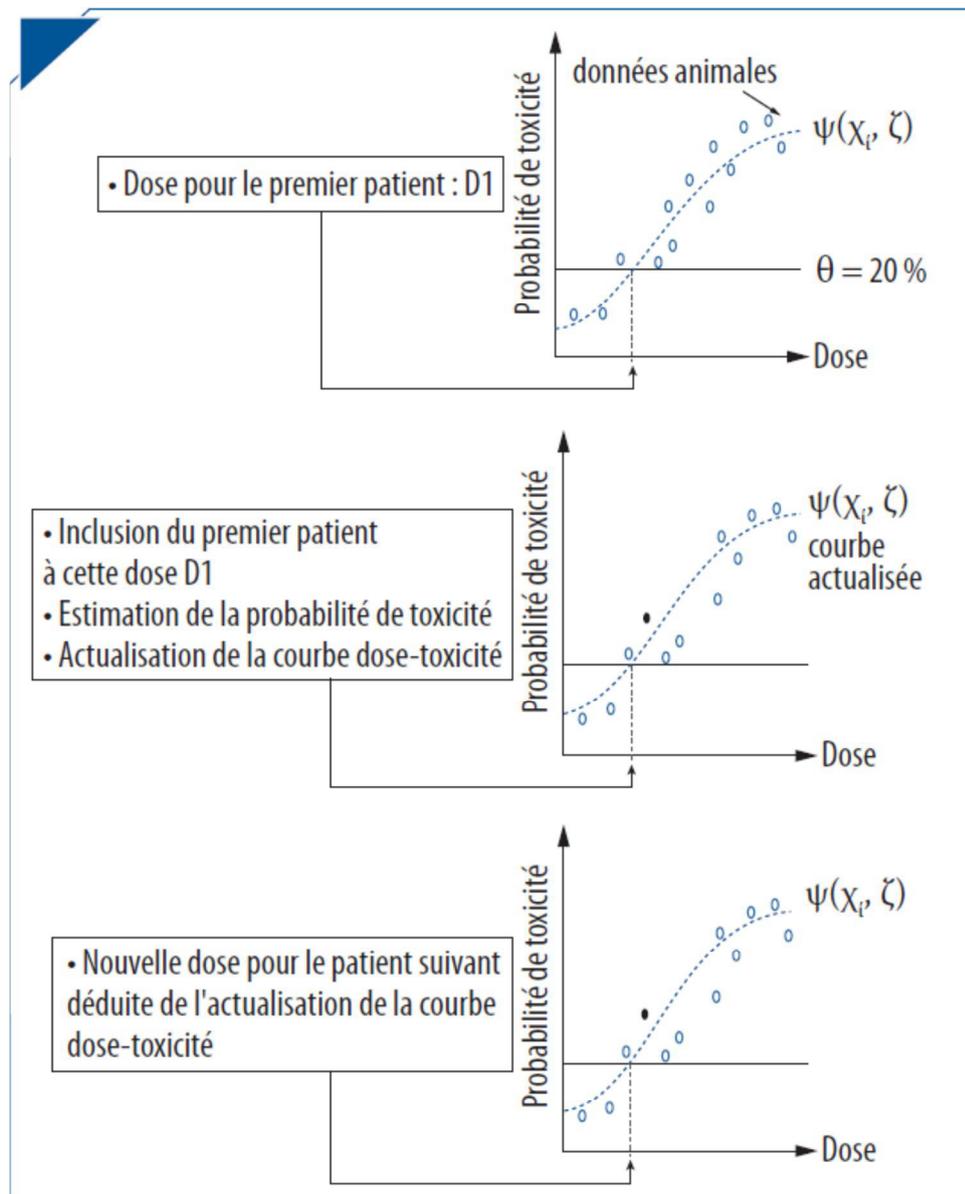
La méthode par réévaluation continue (Continual Reassessment Method ou CRM) est l'approche bayésienne la plus connue dans le contexte des essais de phase I [14]. Cette méthode est basée sur le calcul de probabilité de survenue d'une toxicité à chaque palier de dose, et se déroule en plusieurs étapes (**Figure 3**).

Avant le début de l'essai de phase I, une courbe dose-toxicité est élaborée à partir des données pré-cliniques chez l'animal (en général, à partir d'informations obtenues chez plusieurs animaux différents), mais aussi d'éventuelles données disponibles à partir d'autres essais de phase I. Cette courbe représente la dose en fonction de la probabilité de survenue d'une toxicité, ou bien la fréquence de survenue de toxicité.

Ensuite, à partir de cette courbe dose-toxicité, la dose de départ pour l'essai de phase I est déterminée comme celle correspondant à une probabilité préétablie de toxicité inacceptable (le plus souvent 20%). Le pourcentage peut varier en fonction de la nature de la toxicité et de sa gravité. Un premier patient est alors inclus, à la dose de départ, puis, une fois la toxicité induite définie chez ce patient, la courbe dose-toxicité est réajustée par estimation bayésienne. Ceci permet de déterminer la dose à administrer au patient suivant, correspondant toujours à la probabilité prédéfinie de toxicité inacceptable. Cette procédure continue jusqu'à l'obtention d'une dose stable.

Afin de réduire le risque pour un patient de recevoir une dose provoquant une toxicité inacceptable, plusieurs modifications de cette méthode ont été proposées, portant sur le choix de la dose départ, le nombre de patients affectés par palier de dose et les critères d'arrêt de l'escalade de dose. On peut citer notamment S.N. Goodman et al. [15], qui suggèrent de ne pas se baser sur la distribution « a priori » afin de déterminer le niveau de dose initial, mais plutôt d'utiliser le niveau de dose le plus bas dans l'ensemble de dose ; on parle ainsi de « modified CRM » (ou mCRM). De plus, ils proposent de débiter l'essai de phase I avec des cohortes de 2 ou 3 patients par niveau de dose, et ensuite d'utiliser une méthode type CRM dès l'obtention d'une Toxicité Dose-Limitante (TDL). Enfin, les augmentations de dose sont limitées par des niveaux de dose pré déterminés, ce qui évite de trop grandes escalades de dose.

Figure 3 : Etapes de la méthode CRM : évaluation séquentielle et actualisation



Il est possible d'introduire également dans un essai de phase I la notion d'efficacité, en plus du modèle dose-toxicité traditionnel. Certains auteurs ont ainsi proposé de modifier la CRM en intégrant la durée de suivi des patients ou en réalisant une modélisation simultanée ou séparée des relations dose-efficacité et dose-toxicité. On parle alors de CRM bivariée ou bCRM [16-18].

La méthode CRM (uni ou bivariée) a de nombreux avantages : Premièrement, cette méthode permet d'avoir à chaque instant de l'essai toute l'information préclinique disponible, tout comme celle émanant des autres essais de phase I et même de l'essai de phase I en cours. On réduit ainsi le nombre de patients recevant un traitement à dose infra thérapeutique.

Deuxièmement, du fait de la méthodologie utilisée, la DMT est obtenue avec son intervalle de confiance, et non comme une valeur brute résultant de pourcentages calculés sur 3 ou 6 patients comme avec la méthode selon la suite de Fibonacci. Cette méthode nécessite le soutien de biostatisticiens formés et de techniques de calcul élaborés, mais, désormais, les centres d'oncologie pouvant mettre en place des essais de phase I ont ce soutien à disposition.

Méthode d'escalade guidée par la pharmacocinétique

Dans la majorité des essais de phase I en oncologie, les données de pharmacocinétique-pharmacodynamie (PK-PD) sont mises à l'écart. Ces données PK-PD, issues des essais précliniques, tiennent compte de la variabilité de toxicité et/ou d'efficacité observées entre les patients, et est due en grande partie à des différences dans le devenir du médicament après administration (absorption, distribution dans les différents tissus, métabolisme, élimination). Une analyse PK-PD systématique lors des essais de phase I permettrait, notamment :

- D'identifier les agents avec mécanismes saturables pour lesquels une escalade de dose peut induire une surexposition inattendue au médicament.
- D'identifier d'éventuels métabolites non observés en étude préclinique
- De mettre en évidence d'éventuelles covariables permettant d'identifier des sujets à haut risque de toxicité (par exemple, un patient avec un déficit enzymatique complet en dihydropyrimidine déshydrogénase est exposé à des toxicités majeures en cas d'administration de chimiothérapie type 5-Fluoro uracile)
- De transposer des résultats aux analogues du médicament, en incluant l'information dans de nouveaux essais.

De simuler des essais cliniques afin d'optimiser la méthodologie des essais de phase II ultérieurs (doses et schémas d'administration optimaux).

Concernant la planification de l'escalade de doses, partant de l'hypothèse que des concentrations plasmatiques identiques induisent le même effet biologique chez l'animal et chez l'homme, on procèdera comme précédemment et, au lieu de déterminer une dose maximale tolérée, on cherchera une exposition maximale tolérée, évaluée par l'aire sous la courbe des concentrations plasmatiques (ASC), par une escalade classique, ou selon une CRM en remplaçant la courbe dose/effet par une courbe ASC/effet. Cette approche a été mise en œuvre pour différents cytotoxiques [19-21].

Essais de phase I avec associations de médicaments

L'association de cytotoxiques est une pratique courante en oncologie, permettant le plus souvent d'obtenir un effet synergique sur le plan de l'efficacité. Cependant, ces associations peuvent aussi provoquer des synergies au niveau des toxicités, du fait d'interactions médicamenteuses. Des essais de phase I sont donc nécessaires afin de déterminer la DMT de chaque agent administré au cours d'une même chimiothérapie.

Il existe ainsi un ensemble de plusieurs combinaisons de DMT possibles. L'espace de posologies à administrer est déterminé à partir des résultats d'essais de phase I conduits quand le cytotoxique est étudié en monothérapie. Ensuite, pour l'escalade de dose, plusieurs schémas empiriques existent (exemple : pour deux médicaments, on peut monter la dose du premier en gardant une dose fixe pour le second, escalader les doses alternativement ou de façon simultanée).

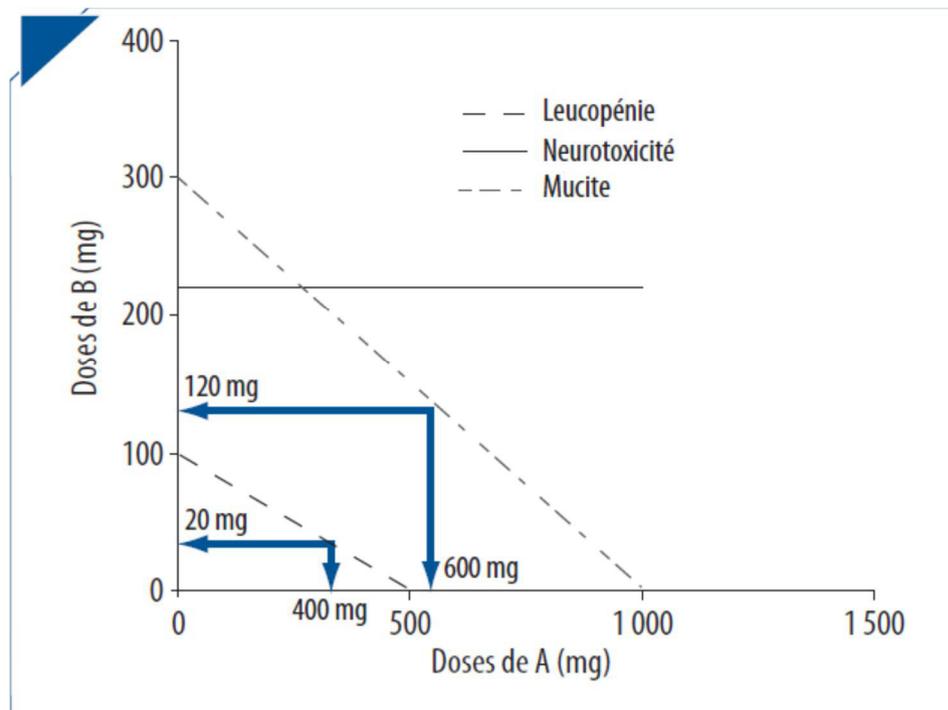
E.L. Korn a élaboré un graphique pour définir l'escalade de doses lors d'associations de deux agents (**Figure 4**) [22]. Un diagramme a été construit à partir de l'information disponible sur chaque médicament pour les différents types de toxicité d'intérêt, en tenant compte de la dose maximale pour chaque agent au regard d'une toxicité donnée et en introduisant un facteur de pondération de toxicité relative de A par rapport à B. Les points situés au-dessus de chaque ligne représentent des combinaisons induisant une toxicité inacceptable. Dans cet exemple, 400 mg de A administrés avec 20 mg de B n'induisent aucun problème d'hémo toxicité. En revanche, si on administre 600 mg de A et 120 mg de B, il faut prévoir des facteurs de croissance tels que le G-CSF pour prévenir la toxicité hématologique ; concernant les autres toxicités, il n'y aura en théorie aucun problème, sauf en cas de synergie de toxicité, celle-ci n'étant pas prise en compte lors de l'utilisation de fonctions linéaires. On peut aussi utiliser cette méthode avec trois agents. Les méthodes les plus sophistiquées en monothérapie, telles que la CRM, peuvent également être appliquées à ces combinaisons [23-24].

Conclusion

La méthode d'escalade de dose selon la suite de Fibonacci modifiée, encore utilisée dans de nombreux essais de phase I en oncologie, est empirique et sans rationnel scientifique. Trop de patients participant à ces essais peuvent être exposés à des doses infra thérapeutiques et laissent de côté une énorme partie de l'information disponible. De plus, le calcul de la DMT manque de précision.

Actuellement, il existe plusieurs nouvelles méthodes, basées sur des concepts scientifiques solides. Elles peuvent être utilisées dans un premier temps pour des essais de phase I d'analogues de molécules bien connues ou pour des essais de phase I avec des médicaments moins toxiques, comme certaines thérapies moléculaires ciblées. Concernant les agents plus toxiques, une méthodologie de type CRM pourrait être privilégiée. Ces informations d'ordre méthodologique sont donc essentielles à connaître lors de la publication d'un article, et leur robustesse doit être vérifiée par le lecteur éclairé.

Figure 4 : Diagramme dose-tolérabilité de deux médicaments A et B en ce qui concerne la leucopénie, la mucite, la neurotoxicité



2-2 Essais de phase II

2-2-1 Objectifs

Au cours du processus de développement d'un nouveau médicament, les essais de phase II jouent un rôle essentiel : Seules les molécules ayant prouvé une efficacité minimale au décours de cette phase, justifieront la poursuite de leur développement dans des essais comparatifs de phase III, essais à plus grande échelle et nécessitant un financement beaucoup plus important. La méthodologie, et plus particulièrement le choix du plan expérimental, de l'essai de phase II doivent donc être méticuleusement planifiés, afin d'éliminer rapidement les traitements inefficaces, tout en minimisant le risque d'écarter une molécule ayant un effet anti tumoral potentiellement intéressant.

Le plus souvent, les essais de phase II sont élaborés à partir de plans expérimentaux à un seul bras, sur un nombre limité de patients, et en utilisant un critère d'évaluation binaire observable à court terme (le plus souvent, la réponse tumorale est utilisée). Cependant, de nouvelles méthodes ont été utilisées ces dernières années, afin de rendre cette phase importante du développement plus efficiente.

2-2-2 Méthodologie

L'essai clinique à un seul bras (bras expérimental) est donc le plan d'expérience le plus utilisé, mais aussi le plus simple. Dans ce type d'essai, le taux de succès observé est comparé à un taux de succès théorique.

Les plans expérimentaux de phase II se basent sur l'hypothèse suivante :

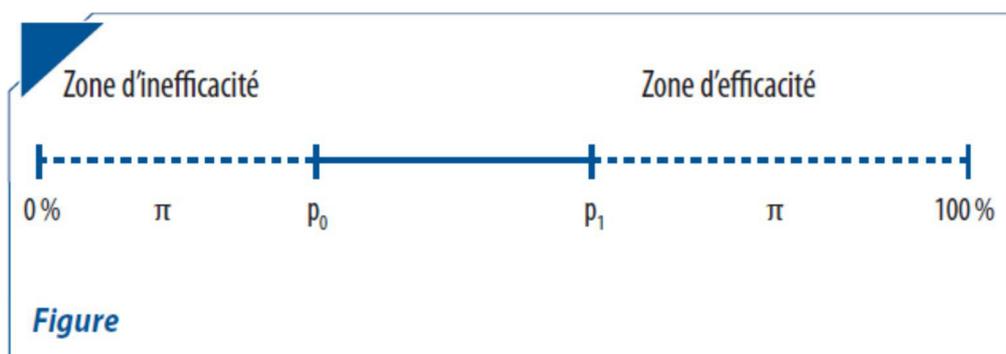
π correspond à la probabilité réelle de succès du traitement à l'étude

- Hypothèse nulle : $H_0 : \pi \leq p_0$
- Hypothèse alternative : $H_1 : \pi \geq p_1$
- p_0 correspond au pourcentage en dessous duquel la molécule sera considérée d'efficacité insuffisante (inefficacité maximale)
- p_1 correspond au pourcentage au-dessus duquel la molécule sera considérée d'efficacité suffisante pour envisager des études ultérieures (efficacité minimale).

Tout l'enjeu de l'essai de phase II va donc consister à rejeter l'hypothèse H_0 et à valider l'hypothèse H_1 .

Les zones d'efficacité et d'inefficacité pour π sont illustrées en **Figure 5**.

Figure 5 : Définition des zones d'efficacité et d'inefficacité pour π



- p_0 et p_1 doivent être clairement définis avant le début de l'étude, ceci est nécessaire pour le calcul du nombre de patients à inclure afin de pouvoir conclure avec une puissance suffisante.
- p_0 correspond le plus souvent à la probabilité de succès du traitement standard, p_1 est supérieur à la probabilité de succès du traitement standard. Pour des raisons éthiques, les plans expérimentaux les plus souvent utilisés sont bâtis en étapes, chacune permettant un arrêt précoce de l'essai en l'absence manifeste d'efficacité, limitant ainsi le nombre de patients traités par une substance inactive. Les plans les plus couramment utilisés sont présentés ci-dessous.

Plan en une étape de Fleming [25]

Il s'agit du plan le plus simple, où le nombre de patients à inclure dans l'essai est défini à l'aide de tables, en fonction des valeurs de p_0 et p_1 fixées a priori, mais aussi du risque α (risque de conclure à l'efficacité d'une molécule qui est en réalité inefficace) et β (risque de ne pas conclure à l'efficacité d'une molécule qui est en réalité efficace). Une seule analyse est réalisée en fin d'essai, où le nombre de patients observés (n_0) est comparé à un nombre de succès théoriques (n). Dans le cas de $n_0 \geq n$, la molécule peut être considérée comme potentiellement efficace. Le risque β doit être minimisé dans un essai de phase II, afin de ne

pas rejeter à tort une molécule potentiellement efficace. β est donc le plus souvent égal à 5% ou 10%. De plus, il n'est pas toujours nécessaire de définir le risque $\alpha=5\%$, comme cela est préconisé dans les essais cliniques de phase III. Avec un risque $\alpha=10\%$, il est nécessaire d'inclure moins de patients, tout en maintenant un risque acceptable d'exposer le patient à une molécule potentiellement inefficace dans un essai de phase III ultérieur.

L'avantage de ce plan d'expérience est sa facilité de mise en œuvre. Par contre, sa limite principale consiste dans le fait qu'aucun arrêt précoce d'étude n'est envisagé, ce qui pose le risque d'exposer un nombre important de patients à une molécule inefficace. Cet inconvénient a été pris en compte ultérieurement lors du développement de plans en plusieurs étapes.

Plans multi étapes: T.R.Fleming [25], E.A.Gehan [26], R.Simon [27], L.G.Ensign et al[28]

Les plans multi étapes utilisent la même méthodologie que dans le plan à une seule étape, mais en y ajoutant des analyses intermédiaires déterminées avant le début de l'essai clinique, afin de vérifier que le nombre de succès obtenu à la fin de chaque étape est supérieur au nombre théorique calculé à priori. Si cette condition est vérifiée à la fin d'une étape, on poursuit jusqu'à l'étape suivante ; sinon la molécule est considérée comme inefficace et l'essai est stoppé définitivement.

La différence entre les plans de T.R.Fleming, E.A.Gehan, R.Simon ou L.G.Ensign réside dans la stratégie d'optimisation du nombre de sujets à inclure lors de chaque étape. Le plan de E.A.Gehan, est optimisé pour arrêter un essai le plus vite possible si aucun succès n'est observé chez les premiers patients inclus. En cas de poursuite, le nombre de patients à inclure dans l'étape 2 tient compte des résultats de la première étape. Pour le plan de T.R.Fleming multi-étapes, il y a une répartition arbitraire du nombre de patients par étape. Ce type de plan permet un arrêt à la fin de l'étape 1 pour inefficacité ou efficacité précoce. En cas d'efficacité précoce, l'essai de phase II est arrêté et la molécule sera jugée comme suffisamment efficace pour être étudiée dans des essais de phase III. R.Simon a développé deux types de plans : le plan « optimum » et le plan « minimax ». Dans le plan « optimum », le nombre de patients exposés à une molécule inefficace est minimisé, tandis que dans le plan « minimax », c'est le nombre total de sujets à inclure dans l'essai qui est minimisé. Enfin, le plan de L.G.Ensign est un plan en trois étapes, combinant la méthode d'E.A.Gehan pour la première étape et celle de R.Simon pour la deuxième et troisième étape.

La principale limite de ces plans multi étapes réside dans le choix du critère de jugement binaire (réponse au traitement par réduction de volume tumoral oui/non), qui n'est pas approprié dans toutes les situations : Il est désormais bien connu que certaines Thérapies Moléculaires Ciblées (TMC) efficaces peuvent entraîner un retard de progression tumorale, voir une augmentation du volume tumorale avec effet de nécrose. Les critères d'évaluation historiques utilisés pour les molécules cytotoxiques, ne sont donc pas applicables dans cette situation, ce qui rend l'utilisation de ces plans difficile. Ceci est aussi vrai avec une nouvelle classe thérapeutique en plein essor, l'immunothérapie, qui peut provoquer des phénomènes de « pseudo progression », avec augmentation sous traitement des volumes cibles, avec un traitement efficace en réalité.

Par ailleurs, le délai nécessaire entre l'inclusion et l'évaluation d'un patient pose parfois problème dans les plans multi-étapes de phase II, avec le dilemme suivant : soit stopper temporairement les inclusions une fois le nombre de patients du palier atteint, au risque d'avoir du mal à reprendre une fois que le passage à l'étape supérieure est effectif, soit prendre le risque que tous les patients soient déjà inclus au moment des résultats de la première étape, ce qui remet en cause l'intérêt de ce type de plan.

Troisièmement, avec ce type de plan, la toxicité du traitement évalué ne peut pas être prise en compte dans les règles de décision. Enfin, du fait de l'absence de bras contrôle, il n'est pas possible de prendre en compte un éventuel biais de sélection.

Essais de phase II randomisés [29-30]

Dans un essai de phase II non randomisé, un traitement est jugé inefficace par confrontation d'un taux de réponse à un taux théorique qui peut être assimilé à une comparaison historique. Or, rien ne permet d'affirmer que le risque de base des patients inclus dans l'essai était comparable à celui des patients ayant permis de déterminer le taux de référence, et ce même si les critères d'inclusion/exclusion semblaient comparables. Ainsi, à l'issue d'un essai de phase II sans bras contrôle, un faible taux de réponse peut être lié à une réelle inefficacité du traitement expérimental, ou bien à la sélection de patients avec une faible probabilité de réponse (biais de sélection). L'ajout d'un bras de référence jouant le rôle de bras contrôle et la répartition des patients entre traitement expérimental et traitement de référence par tirage au sort permettent de limiter les effets de ce biais de sélection. Le taux de réponse observé dans le bras traitement de référence sert ainsi à valider la mesure de celui observé dans le bras expérimental. Contrairement à un essai de phase III, ce type de plan expérimental n'est en aucun cas prévu pour comparer les effets de traitements entre les deux bras, il sert uniquement à renforcer la valeur des conclusions sur l'efficacité potentielle du traitement expérimental si, à l'issue de l'essai, le taux de réponse observé dans le bras traitement de référence s'avère proche de celui attendu.

Le calcul du nombre de sujets est réalisé, pour le bras testé, de la même façon que pour un essai de phase II classique. Le nombre de patients à inclure dans le bras contrôle est fixe de manière que l'effectif des deux bras soit égal. On peut également envisager de mettre moins de patients dans le bras de référence (randomisation 1:2), qui sert seulement à vérifier l'hypothèse p_0 avec une précision correcte.

Essais avec prise en compte de la toxicité

Dans les essais de phase II classiques, la toxicité de la molécule n'est pas prise en compte dans les règles de décision. Ils existent plusieurs plans d'expérience pour pallier à ce problème. Ainsi, il est ajouté au critère de jugement principal (efficacité) un critère secondaire permettant de juger de la toxicité. Des règles d'arrêt combinant taux de réponse et taux de toxicité sont ensuite définies.

Voici un exemple [31] :

L'essai de phase II est basé sur un plan d'expérience à n étapes, conçu pour évaluer la chimiothérapie forte dose dans le traitement des lymphomes malins non hodgkiniens. Le traitement est jugé prometteur si son taux de réponse (R) est supérieur à celui du traitement standard (R_0) et que son taux de toxicité (T) est inférieur à celui du traitement standard.

Les hypothèses H_0 et H_1 peuvent alors être écrites sous la forme suivante :

$H_0 : PR \leq PR_0$ ou $PT \geq PT_0$ et $H_1 : PR > PR_0$ et $PT < PT_0$.

Les hypothèses peuvent être représentées comme sur la **Figure 6**, avec les probabilités d'événements définies à priori suivantes :

- PR_0 : Probabilité de réponse du traitement standard : 0.5
- PT_0 : Probabilité de toxicité du traitement standard : 0.3
- $PT_0 \mid PR_0$: Probabilité de réponse en cas de toxicité : 0.35
- PR : Probabilité de réponse attendue pour le nouveau traitement : 0.75
- PT : Probabilité de toxicité attendue pour le nouveau traitement : 0.15

Les risques de première et deuxième espèces sont définis de la manière suivante :

$\alpha_R : (PR \leq PR_0 \text{ et } PT \geq PT_0) \leq 5 \%$

$\alpha_T : (PR \leq PR_0 \text{ ou } PT \geq PT_0) \leq 30 \%$

$1 - \beta : (PR > PR_0 \text{ et } PT < PT_0) > 90 \%$

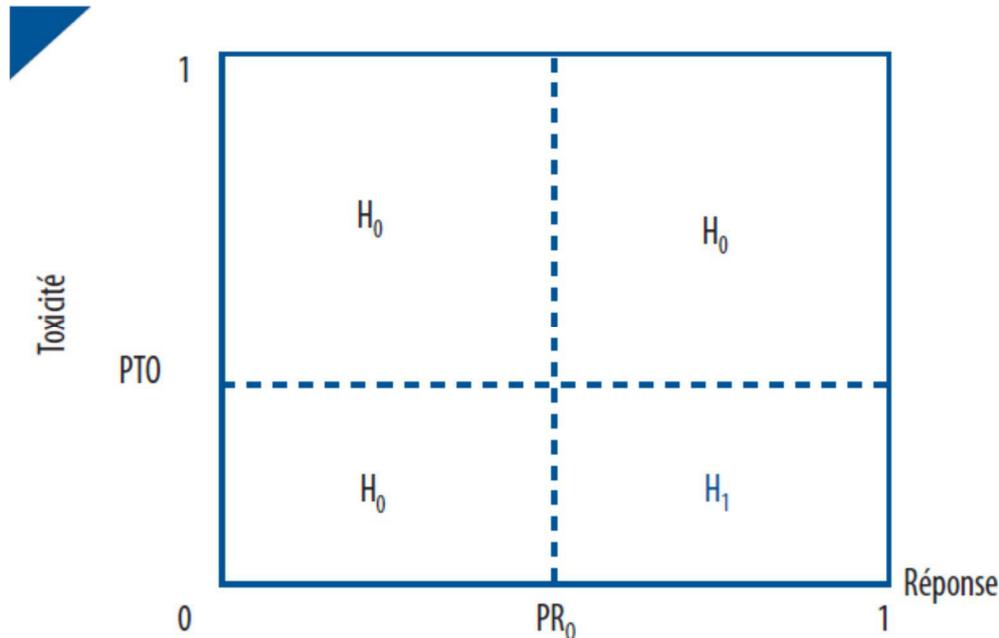
Dans cet exemple, les auteurs utilisent un algorithme qui permet de minimiser la taille de l'essai et de déterminer les règles d'arrêt à chacune des n étapes.

Plan bayésien [32-34]

L'approche séquentielle, ou bayésienne, est de plus en plus utilisée pour les essais de phase II en cancérologie. Avec ce type de plan, l'hypothèse de départ peut se modifier en fonction des données recueillies au cours de l'essai. Des analyses intermédiaires sont ainsi réalisées lors de l'étude, tous les n patients. Chacune de ces analyses fait évoluer la loi de probabilité du critère principal en se basant sur les informations concernant les premiers patients. La réponse au traitement au début de l'essai est considérée comme une variable ayant une distribution de probabilité $P(\theta)$ [*prior*] peu discriminante. Au fil de l'essai, les données observées permettent d'actualiser $P(\theta)$, afin d'obtenir une distribution de probabilité à postériori (*posterior*) de la réponse au traitement, qui devient de plus en plus informative au fur et à mesure de l'avancement de l'essai.

Le nombre de sujets à inclure n'est pas fixé avant le début de l'essai, mais va dépendre des données observées au cours de l'étude. Ces types de plans ont l'avantage de limiter le nombre de patients à inclure, d'autant plus si le traitement est efficace.

Figure 6 : Représentation des hypothèses H_0 et H_1 de l'exemple3 : PR_0 : Probabilité de réponse du traitement standard : 0.5 ; PT_0 : Probabilité de toxicité du traitement standard : 0.3 ; $PT_0 | PR_0$: Probabilité de réponse en cas de toxicité : 0.35



Les limites de ce type de plan sont dues au fait que le *prior* est difficile à définir, ou que le *posterior* devient trop rapidement informatif, avec des conclusions pouvant être tirées trop hâtivement (par exemple, arrêt de l'étude si premier patient en échec). Ce type de limitation peut être limité en définissant un nombre minimal de patients à inclure. Les études utilisant un plan bayésien nécessitent une grande réactivité dans le monitoring des données, et ceci peut être d'autant plus difficile si l'essai est multicentrique.

Essais de phase II-III [35-36]

Lors du développement d'une nouvelle molécule anticancéreuse, il peut parfois s'écouler un certain laps de temps entre la fin d'un essai de phase II ayant démontré une efficacité anti tumorale, et l'initiation de l'essai de phase III ultérieur. Les essais de phase II-III ont pour but de réduire cet intervalle de temps entre essais, en utilisant les données des patients inclus dans l'essai de phase II pour l'essai de phase III. Cette approche est possible en utilisant les critères d'arrêt précoce issus des essais de phase II dans des essais de phase III, ce qui pourrait être assimilable au principe d'une analyse intermédiaire. L'avantage de ce type de plan est de réaliser un essai de phase II avec n bras expérimentaux, puis de sélectionner le plus efficace comme bras expérimental pour l'essai comparatif de phase III.

Les essais de phase II-III n'ont d'intérêt que si les critères d'efficacité utilisés pour les phases II et III sont fortement corrélés, voire similaires, ce qui n'est pas toujours le cas. En effet, en oncologie, la réponse tumorale est souvent utilisée comme critère d'efficacité dans les essais de phase II, puis ce critère est par la suite abandonné au profit de la survie globale ou de la survie sans progression dans l'essai de phase III.

Conclusion

Les essais de phase II occupent une place très importante dans le processus de développement d'un agent anticancéreux. Les plans expérimentaux sont multiples, le choix du meilleur plan devra tenir compte du potentiel de recrutement et du laps de temps nécessaire pour ce recrutement. Il faut être particulièrement attentif à la définition des risques d'erreurs α et β , la puissance doit être priorisée afin de ne pas écarter une molécule potentiellement efficace. Il est également nécessaire de retenir que l'ajout d'une randomisation avec bras contrôle de référence dans un essai de phase II ne permet en aucun cas de comparer directement bras expérimental et bras contrôle avec une puissance suffisante.

Tout comme pour les essais de phase I, ces informations d'ordre méthodologique doivent être rapportées clairement lors de la publication d'un essai de phase II, afin de pouvoir juger de la qualité et de la précision des résultats obtenus.

2-3 Essais de phase III

2-3-1 Objectifs

L'essai de phase III est l'étude comparative d'efficacité proprement dite. Son but est de tester l'efficacité d'un nouveau traitement par rapport au meilleur traitement connu, également appelé traitement de référence. Au cours de cette phase, l'efficacité et la tolérance sont estimées en situation réelle. Les essais de phase III en oncologie sont des essais de grande envergure, pouvant impliquer plusieurs centaines à plusieurs milliers de patients. Les durées d'exposition au médicament testé sont plus longues que dans les essais de phase II, ce qui permet notamment de réaliser des analyses d'interactions médicamenteuses. Ce n'est seulement qu'à l'issue de cette phase, qu'une Autorisation de Mise sur le Marché (AMM) pourra être donnée par les autorités de santé.

Les objectifs des essais de phase III sont bien définis et la méthodologie utilisée est basée essentiellement sur la minimisation des biais. Cela conduit à réaliser des essais randomisés, contrôlés prospectifs, comparatifs, idéalement en double aveugle, avec un critère de jugement principal unique et défini clairement, ainsi qu'à une analyse statistique réalisée en intention de traiter.

2-3-2 Randomisation

Le principe de randomisation est obligatoire dans un essai de phase III : Il consiste à tirer au sort, à l'aide d'une procédure centralisée et imprévisible pour l'investigateur, le traitement administré (de référence ou expérimental) pour chaque nouveau patient inclus dans l'essai. En effet, le but de cette phase étant de comparer directement les traitements testés

entre eux, il convient de s'assurer que la constitution des groupes est réalisée sans biais. La randomisation permet d'assurer cette attribution de manière aléatoire et d'éviter ainsi un biais de sélection.

L'investigateur et le patient peuvent avoir différents niveaux de connaissance du traitement administré, permettant d'exclure également totalement les biais au niveau de leur estimation de l'effet du traitement :

- Essai en double aveugle : la nature du traitement alloué par randomisation est inconnue du médecin et du patient (essai à fort niveau de preuve).
- Essai en simple aveugle : la nature du traitement alloué par randomisation est connue soit par le médecin, soit par le patient (essai à niveau de preuve intermédiaire).
- Essai en ouvert : la nature du traitement alloué par randomisation est connue du médecin et du patient (niveau de preuve plus faible).

En cancérologie, le double aveugle est souvent impossible à respecter : Les traitements testés ont souvent des effets secondaires caractéristiques, ce qui permet à l'investigateur de savoir pour son patient dans quel bras de traitement il a été randomisé. Il est possible de compenser cette limite en utilisant un critère de jugement non subjectif, comme la mortalité, ou de faire des essais en PROBE : c'est-à-dire que l'évaluation du critère de jugement se fait à l'aveugle.

2-3-3 Critère principal

Il s'agit du critère unique sur lequel va se baser la conclusion de l'essai. C'est sur ce critère que le nombre de patients à inclure dans l'essai est calculé et c'est donc lui seul qui garantit la puissance requise pour pouvoir conclure. Contrairement aux essais de phase II, pour lesquels le critère de jugement est souvent un critère binaire rapidement observable, les essais de phase III utilisent un critère à moyen ou long terme. Il s'agit préférentiellement d'un critère clinique, le plus objectif possible, et non intermédiaire. Ce critère doit être défini le plus précisément possible, et doit correspondre au critère clinique le plus pertinent vis-à-vis de l'objectif de l'étude.

En oncologie, le critère principal est souvent défini comme la survie globale ou la survie sans progression. En plus du critère de jugement principal, d'autres facteurs d'intérêt peuvent être analysés : Il s'agit des critères secondaires. Cependant, pour ces critères, la puissance statistique n'est pas garantie et peut donc être faible

2-3-4 Types d'essais

L'objectif de l'essai de phase III peut être de trois ordres :

- Montrer la supériorité du traitement expérimental par rapport au traitement de référence (essai de supériorité ou d'efficacité).

- Montrer l'équivalence du traitement expérimental par rapport au traitement de référence (essai d'équivalence).
- Montrer la non-infériorité du traitement expérimental par rapport au traitement de référence (essai de non-infériorité).

Essai de supériorité ou d'efficacité

Il s'agit de la majorité des essais de phase III en oncologie. Les hypothèses sont définies de la manière suivante :

- H_0 : hypothèse nulle, à rejeter : Le traitement expérimental est inférieur ou égal au traitement de référence
- H_1 : hypothèse alternative : Le traitement expérimental est supérieur au traitement de référence

Lorsque l'hypothèse H_0 ne peut pas être rejetée, l'essai est non concluant. L'absence de supériorité ne permet pas de conclure à une équivalence ou à une non-infériorité.

Essai d'équivalence ou de non-infériorité

Ce type d'essai a souvent comme objectif de comparer deux formes galéniques, deux rythmes d'administration ou bien deux molécules appartenant à la même classe thérapeutique. Ici, le but n'est pas de prouver l'efficacité supérieure du traitement expérimental, mais de démontrer une plus grande facilité d'utilisation, une meilleure tolérance globale ou un coût moindre.

Sur le plan statistique, il est donc nécessaire de définir à priori une zone d'équivalence. La valeur-seuil de cette zone correspond à la perte d'efficacité maximale acceptée avec le traitement expérimental. Le choix de cette valeur seuil est donc primordial. Plus cette zone d'équivalence est réduite, plus le nombre de patients à inclure sera grand.

Les hypothèses sont donc les suivantes :

- En situation d'équivalence (test bilatéral):

H_0 : les traitements ne sont pas équivalents

H_1 : les traitements sont équivalents

- En situation de non-infériorité (test unilatéral):

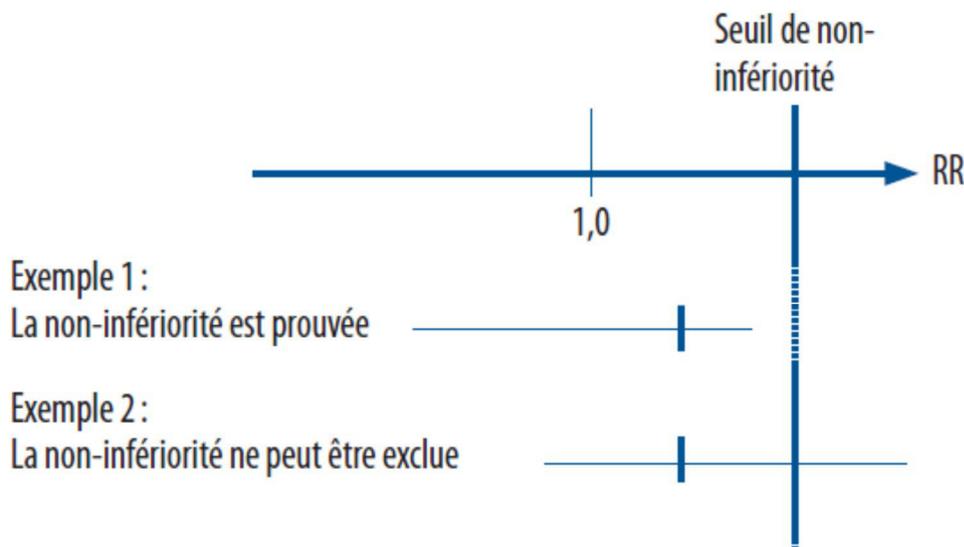
H_0 : le nouveau traitement est inférieur au traitement de référence

H_1 : le nouveau traitement est égal ou supérieur au traitement de référence

Le test bilatéral est peu utilisé car il existe un risque de rejeter l'équivalence si le traitement expérimental est supérieur au traitement de référence. L'essai de non-infériorité est en effet plus intéressant pour le clinicien, il s'agit ici de prouver que le traitement à l'étude est au moins aussi efficace que le traitement de référence.

La méthodologie statistique utilisée dans ce type d'essai consiste à créer un intervalle de confiance autour du risque relatif, et à comparer la borne supérieure de cet intervalle avec la limite de non-infériorité définie a priori (**Figure 7**).

Figure 7 : Essai de non-infériorité



2-3-5 Schéma expérimental

La priorité dans les essais de phase III est de limiter l'exposition aux biais. Les différentes méthodes permettant la réduction des biais dans les essais contrôlés randomisés sont les suivantes :

- Biais de sélection : Recours à un essai randomisé, tirage au sort non prévisible
- Biais d'attrition : Réalisation d'analyse en intention de traiter, rigueur de suivi pour éviter les perdus de vue, méthodes d'analyses des données manquantes
- Biais de suivi : Recours à un essai en double aveugle
- Biais d'évaluation : Recours à un essai en double aveugle
- Biais de confusion : Réalisation d'un essai comparatif ou contrôlé

2-3-6 Conclusion

Une méthodologie importante est disponible pour les essais de phase III, pour aboutir à une conclusion non biaisée de l'effet du traitement. Le budget mobilisé pour un essai de phase III étant conséquent, il est primordial d'aboutir au décours de l'essai à une conclusion fiable sur l'efficacité du nouveau traitement.

Ces informations d'ordre méthodologiques doivent donc être clairement rapportées et accessibles lors de la publication des résultats de ces essais, afin de pouvoir évaluer la validité et la pertinence de la méthodologie mise en œuvre pour répondre à la question posée initialement.

La méthodologie des essais cliniques lors du développement des médicaments a changé depuis plusieurs années, notamment avec l'application de méthodes bayésiennes et l'apparition de nouveaux schémas (essais adaptatifs, essais en clusters). L'essor de nouvelles classes thérapeutiques en oncologie, telles que l'immunothérapie ou la Thérapie Moléculaire Ciblée (TMC), a provoqué un intérêt croissant des différentes firmes pharmaceutiques pour le développement de nouvelles molécules, ceci provoquant une augmentation importante du nombre d'essais ouverts en oncologie (essais de phase I, II, III). Plus récemment, l'élaboration d'essai d'associations chimiothérapie + immunothérapie, chimiothérapie + TMC, chimiothérapie + immunothérapie+ TMC a complexifié les stratégies thérapeutiques. Une méthodologie précise et de qualité de ces essais cliniques reste, plus que jamais, un prérequis indispensable afin de pouvoir tirer une conclusion fiable et fondée sur les faits. La description détaillée d'éléments tels que le schéma expérimental, le choix et la définition du critère de jugement principal, le calcul du nombre de sujets nécessaires sont essentielles pour juger de la validité et de la reproductibilité de l'essai.

Avant la publication d'un essai clinique dans un journal médical, le comité de relecture vérifie préalablement la qualité et la transparence des informations (notamment d'ordre méthodologique), définies sous le terme générique de reporting.

3- Le reporting des études et son évaluation

3-1 Histoire du reporting

Depuis plusieurs décennies, il est établi que les essais cliniques randomisés, avec un design approprié à la question pré établie, et qui se réalisent conformément au protocole planifié a priori, permettent de fournir les résultats les plus robustes sur l'efficacité d'une intervention thérapeutique. Cependant, les essais cliniques qui ont une méthodologie inadaptée sont très souvent associés à des biais, notamment avec une exagération des effets de traitement [37-40]. Des résultats biaisés, issus d'un essai clinique avec un design inadapté peuvent avoir un impact néfaste sur les décisions médicales prises, ceci pouvant aller du choix de traitement inadapté pour les patients à l'élaboration de recommandations de prise en charge thérapeutique reposant sur ces études biaisées.

L'analyse critique de la qualité d'un essai clinique n'est possible que si son schéma expérimental, sa réalisation et ses conditions d'analyse des résultats sont clairement et précisément décrits dans l'article publié. Le reporting est le fait de décrire dans une publication toutes les actions, notamment d'ordre méthodologiques, réalisées au cours de l'étude.

Entre 1990 et 2000, plusieurs études ont démontré que le reporting des essais cliniques est souvent incomplet et manque de transparence [41-44].

3-1-1 Un reporting incomplet et peu précis des essais cliniques

De nombreuses revues de littérature ont mis en lumière les déficiences dans le reporting des essais cliniques. Par exemple, les informations sur la méthode utilisée pour assigner un participant au groupe de comparaison dans les essais randomisés n'était rapporté

que dans seulement 21% des 519 rapports d'essais indexés dans PUBMED en 2000 [45] et dans seulement 34% des 616 rapports indexés en 2006 [46]. De même, seulement 45% des rapports d'essais indexés dans PUBMED en 2000 [45] et 53% en 2006 [46] avaient clairement défini le critère de jugement principal, et seulement 27% en 2000 et 45% en 2006 avaient rapporté le calcul de la taille d'échantillon nécessaire à l'essai. Il a été également démontré que les rapports d'essais cliniques sont non seulement souvent incomplets, mais parfois aussi inexacts : Sur 119 rapports d'essais indiquant que tous les participants ont été inclus dans l'analyse, dans les groupes auxquels ils ont été assignés à l'origine (analyse en intention de traiter), 15 (13%) ont exclus des patients ou n'ont pas analysé tous les patients comme prévu [47]. De nombreuses autres revues de littérature ont montré que le reporting incomplet d'un essai est courant dans les revues spécialisées [45-48] et les revues publiées dans des langues autres que l'anglais [49-50].

DerSimonian et ses associés ont suggéré que “les éditeurs pouvaient améliorer considérablement le reporting des essais cliniques en fournissant aux auteurs une liste d'items qui devraient être strictement rapportés dans le manuscrit. ” [51] Au début des années 90, deux groupes d'éditeurs de journaux, de promoteurs d'essais cliniques et de méthodologistes ont publié indépendamment des recommandations sur le reporting des essais [52-53].

Dans un éditorial ultérieur, Rennie a exhorté les deux groupes à se rencontrer et à élaborer des recommandations communes [54]. Ces rencontres ont finalement abouti à la publication de la déclaration CONSORT (Consolidated Standards of Reporting Trials) [55].

3-1-2 La déclaration CONSORT

La première déclaration CONSORT comprenait une liste d'items essentiels à inclure dans les rapports d'essais cliniques randomisés (**Figure 8**), ainsi que des explications pour élaborer le diagramme de flux des patients inclus [55]. Initialement, ces recommandations ciblaient surtout le reporting des essais cliniques randomisés à deux groupes, selon un plan d'expérience en groupes parallèles. Cependant, une grande partie de la checklist CONSORT était aussi applicable à d'autres types de design d'essais, tels que les essais de non infériorité, d'équivalence, les essais en plan factoriel ou en plan croisé. Il existe des extensions à la déclaration CONSORT, portant spécifiquement sur les designs cités ci-dessus, dont certaines ont été publiées [56-58]. D'autres extensions publiées se sont focalisées sur le reporting de données particulières, telles que celles portant sur les dommages dans les essais cliniques [59], sur les traitements non pharmacologiques [60], sur les traitements à base de plantes [61], sur les abstracts [62].

L'objectif principal de la déclaration CONSORT consistait à guider les auteurs, pour améliorer la qualité de reporting de leurs essais. Les manuscrits d'essais cliniques, publiés dans des journaux médicaux, ont en effet besoin d'être clairs, transparents et complets. Les lecteurs, les comités de relecture ainsi que les éditeurs, pouvaient utiliser CONSORT, pour les aider à évaluer de manière critique et à interpréter les rapports d'essais cliniques. Le contenu

de la checklist CONSORT est axé principalement sur des points relatifs à la validité interne et externe des essais cliniques.

Figure 8 : La liste des items de CONSORT (version publiée en 1996)

Heading	Subheading	Descriptor	Was It Reported?	On What Page No.?
Title		Identify the study as a randomized trial. ⁷		
Abstract		Use a structured format. ^{8,9}		
Introduction		State prospectively defined hypothesis, clinical objectives, and planned subgroup or covariate analyses. ¹⁰		
Methods				
	Protocol	Describe Planned study population, together with inclusion/exclusion criteria. Planned interventions and their timing. Primary and secondary outcome measure(s) and the minimum important difference(s), and indicate how the target sample size was projected. ^{2,11} Rationale and methods for statistical analyses, detailing main comparative analyses and whether they were completed on an intention-to-treat basis. ^{12,13} Prospectively defined stopping rules (if warranted). ¹⁴		
	Assignment	Describe Unit of randomization (eg, individual, cluster, geographic). ¹⁵ Method used to generate the allocation schedule. ¹⁶ Method of allocation concealment and timing of assignment. ¹⁷ Method to separate the generator from the executor of assignment. ^{17,18}		
	Masking (Blinding)	Describe mechanism (eg, capsules, tablets); similarity of treatment characteristics (eg, appearance, taste); allocation schedule control (location of code during trial and when broken); and evidence for successful blinding among participants, person doing intervention, outcome assessors, and data analysts. ^{19,20}		
Results				
	Participant Flow and Follow-up	Provide a trial profile (Figure) summarizing participant flow, numbers and timing of randomization assignment, interventions, and measurements for each randomized group. ^{3,21}		
	Analysis	State estimated effect of intervention on primary and secondary outcome measures, including a point estimate and measure of precision (confidence interval). ^{22,23} State results in absolute numbers when feasible (eg, 10/20, not 50%). Present summary data and appropriate descriptive and inferential statistics in sufficient detail to permit alternative analyses and replication. ²⁴ Describe prognostic variables by treatment group and any attempt to adjust for them. ²⁵ Describe protocol deviations from the study as planned, together with the reasons.		
Comment		State specific interpretation of study findings, including sources of bias and imprecision (internal validity) and discussion of external validity, including appropriate quantitative measures when possible. State general interpretation of the data in light of the totality of the available evidence.		

Il est important de noter que la déclaration CONSORT n'a pas été conçue comme un outil d'évaluation de la qualité de l'étude.

Depuis sa publication en 1996, CONSORT a été soutenu par plus de 400 revues (www.consort-statement.org) et par plusieurs groupes de rédaction, tels que le comité international des éditeurs de revues médicales [63]. Depuis la publication de la première version, la déclaration CONSORT est révisée périodiquement : La première révision majeure a été publiée en 2001 [64], puis la deuxième en 2010 [65].

3-1-3 La déclaration STROBE

A l'instar des essais cliniques, de nombreux travaux ont démontré qu'il existait des lacunes dans le reporting des études observationnelles. Pocock SJ et al [66] ont réalisé une analyse des études épidémiologiques, publiées dans des revues médicales généralistes et spécialisées, et ils ont constaté que le rationnel expliquant le choix des potentielles variables de confusion était souvent non rapporté. Concernant les publications d'études cas-témoins en psychiatrie, il a aussi été observé un défaut du reporting sur la méthode utilisée pour définir les cas et les témoins [67]. Dans une enquête sur les études longitudinales dans la recherche sur les AVC, 17 articles sur 49 (35%) ne précisaient pas les critères d'éligibilité [68].

Certains auteurs ont ainsi déclaré qu'il était nécessaire d'élaborer des recommandations portant sur le reporting des études observationnelles [69-70]. Un groupe de travail a donc été créé en 2004, comprenant des méthodologistes, des chercheurs et des éditeurs afin d'élaborer les recommandations sur le reporting des études observationnelles : La déclaration STROBE (Strengthening the Reporting of Observational Studies in

Epidemiology) [71]. Etant donné que la recherche observationnelle regroupe de nombreux schémas expérimentaux d'études différents, le champ d'application des recommandations STROBE a été défini dès le début du travail. Ainsi, ces recommandations se sont concentrées sur les trois modèles d'études les plus utilisés en recherche analytique : Les cohortes, les études cas-témoins et les études transversales. La déclaration STROBE comprenait une liste de 22 items qui devaient être rapportées dans les publications de ces études. L'objectif des auteurs de STROBE consistait à aider les cliniciens et les chercheurs à rapporter de manière claire et exhaustive leur recherche observationnelle. Ces recommandations ne devaient en aucun cas être considérées comme des instructions dans la conception ou dans le déroulement d'une étude observationnelle.

Tout comme pour CONSORT, il a été ensuite publié de nombreuses extensions à la déclaration STROBE, pour, entre autres, les études d'associations dans les maladies génétiques [72], les études d'épidémiologie moléculaire en infectiologie [73], les études épidémiologiques portant sur la résistance aux traitements antimicrobiens [74].

L'ensemble des publications (recommandations et extensions de recommandations) sont accessibles via le site de l'EQUATOR network (Enhancing the QUALity and Transparency Of health Research) [75].

3-1-4 La déclaration PRISMA

Actuellement, il est reconnu que les revues systématiques de la littérature et les méta-analyses bien réalisées ont une valeur scientifique importante dans le domaine de la recherche médicale. Les médecins cliniciens lisent ces études afin de se mettre à jour dans leur domaine de compétence [76-77] et ces dernières servent souvent de socle pour l'élaboration et la mise à jour de recommandations de bonne pratique clinique.

Comme pour toute recherche, la valeur d'une revue systématique ou d'une méta-analyse dépend de ce qui a été fait, de ce qui a été trouvé et de la clarté de son reporting. Cette qualité du reporting est variable, limitant la capacité des lecteurs à évaluer les forces et faiblesses de ces études.

Plusieurs études avaient déjà évalué plus spécifiquement la qualité du reporting des revues systématiques. En 1987, Mulrow et al avaient analysé 50 revues, publiées dans quatre journaux médicaux en 1985 et 1986, et ils avaient constaté qu'aucune revue ne rapportait complètement un ensemble de huit critères méthodologiques majeurs, tels que l'évaluation de la qualité intrinsèque des études prises en compte dans la revue de littérature [78]. En 1987, Sacks et al. [79] avaient évalué le contenu du reporting de 83 méta-analyses, sur 23 caractéristiques et dans six domaines distincts. La qualité du reporting était généralement médiocre : Entre une et 14 de ces caractéristiques étaient adéquatement rapportées (moyenne = 7,7 ; écart type = 2,7). La mise à jour de cette étude en 1996 avait démontré l'absence de progrès dans la qualité du reporting [80].

En 1999, afin de remédier à la qualité sous-optimale du reporting des méta-analyses, un groupe international a élaboré des premières recommandations, appelées déclaration QUOROM (Quality Of Reporting Of Meta-analyses), portant exclusivement sur le reporting

de méta-analyses d'essais contrôlés randomisés [81]. En 2009, ces recommandations ont fait l'objet d'une mise à jour, notamment en y intégrant les revues systématiques de littérature : La déclaration QUOROM a ainsi été renommée PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) [82].

La déclaration PRISMA contenait une liste de 27 items, ainsi que 4 diagrammes de flux. L'objectif de ces recommandations consistait à aider les auteurs à améliorer la qualité de reporting des méta-analyses et des revues systématiques publiées. Lors de la publication, les auteurs de PRISMA avaient également précisé que ces recommandations ne pouvaient pas être considérées comme un outil d'évaluation de la valeur scientifique globale d'une revue systématique ou d'une méta-analyse.

Par la suite, plusieurs extensions de PRISMA ont été publiées, portant notamment sur le reporting des toxicités/effets secondaires des traitements dans les revues systématiques [83], sur le reporting des revues systématiques et méta-analyses à données individuelles [84], ou sur le reporting de revues systématiques et de méta-analyses portant sur les études de test de performance diagnostique [85]. Comme la recommandation STROBE et ses extensions, ces publications sont en libre accès via le site de l'EQUATOR network [75].

3-1-5 Les autres déclarations

Il existe dans la littérature de nombreuses autres recommandations portant sur le reporting des essais, interventionnels ou non, et prises en compte par les comités de relecture des journaux médicaux.

Parmi ces dernières, la déclaration REMARK (Reporting Recommendations for Tumor Marker Prognostic Studies) [86] a été élaborée comme support, applicable de la conception initiale à l'analyse finale des études de marqueur tumoraux et des études médicales pronostiques dans un registre plus général. Lors de l'élaboration de REMARK, les auteurs se sont appuyés sur de nombreuses publications, constatant que la littérature disponible portant sur des marqueurs pronostiques en cancérologie (et dans d'autres spécialités) présentait de nombreuses lacunes, d'ordre méthodologique et en termes de reporting [87-89].

La déclaration REMARK comprenait une liste de 20 items, ainsi qu'un diagramme de flux des patients, avec un focus particulier sur certains éléments comme la description du matériel biologique utilisé (type de matériel, conditions de conservation et de stockage), la méthodologie d'analyse des marqueurs biologiques (protocoles, kits et réactifs utilisés, qualité des procédures de contrôle, reproductibilité des analyses, méthodes de quantification, protocole de notation), le design de l'étude (notamment la méthodologie de sélection des cas, la description complète des variables examinées initialement ou envisagées pour inclusion dans le modèle) et l'analyse des résultats. Pour chaque item, les auteurs avaient ajouté un ou plusieurs exemples de reporting adéquat, issus de publications.

La déclaration CARE (CAse Report) est dédiée au reporting des rapports de cas [90]. Ce type d'étude a déjà prouvé son importance notamment dans la description de maladies rares ou de nouvelles maladies, dans l'évaluation des effets bénéfiques ou délétères d'une intervention, mais aussi en pédagogie médicale [91-92]. La checklist CARE comprend 13 items de reporting, afin d'aider les auteurs à reporter précisément leurs rapports de cas.

De nombreuses autres déclarations de reporting ont été publiées ces dernières années [93-95], portant sur des types d'études très variées (études médico-économiques : CHEERS, reporting de protocoles d'études, études qualitatives : COREQ, etc). L'ensemble de ces recommandations sont accessibles via le site de l'EQUATOR network [75].

3-2 Evaluation de la qualité de reporting des études

Depuis leurs publications, les différentes recommandations de reporting (CONSORT, STROBE, PRISMA, REMARK, etc) ont été utilisées par les auteurs comme des outils précieux d'aide à la rédaction des manuscrits, afin de garantir une clarté et une transparence optimale de l'information disponible pour le lecteur lors de la publication. D'un autre côté, les différents éditeurs de journaux ont pu prendre en compte ces déclarations et les utiliser afin d'évaluer la qualité de reporting des différents articles soumis. Les comités de relecture des journaux médicaux les plus prestigieux intègrent désormais les déclarations de reporting dans leur lecture critique des différentes études en attente de publication.

Cependant, peut-on affirmer que la qualité du reporting des études biomédicales s'est améliorée depuis la publication de ces différentes déclarations ? Existe-t-il des spécialités médicales ou paramédicales particulières où la qualité du reporting des études a été plus analysée ?

Afin de répondre à ces questions, une revue de littérature narrative a été réalisée, via la base de données Medline (par Pubmed), afin d'identifier l'ensemble des publications dont le sujet portait sur l'évaluation de la qualité du reporting des études biomédicales. La dernière recherche a été actualisée en Février 2019, avec la stratégie de recherche suivante : Reporting[Title/Abstract] AND Quality[Title/Abstract]) AND statement[Title/Abstract].

Afin d'être éligibles, les études devaient être publiées entre Juillet 2009 et Février 2019, correspond ainsi au laps de temps écoulé depuis la publication officielle de la déclaration PRISMA[81], le type d'étude devait être une revue systématique de littérature ou une étude transversale, l'objectif principal de l'étude devait consister à évaluer la qualité du reporting d'essais interventionnels ou non interventionnels, préalablement publiés, à l'aide d'une ou plusieurs recommandations de reporting. Le langage d'écriture devait être en anglais, au moins pour la rédaction de l'abstract.

Il n'a pas été pris en compte les lettres à l'éditeur, ainsi que les revues qui n'utilisaient pas les déclarations de reporting, publiées et accessible via l'EQUATOR network, comme outil d'évaluation.

Parmi toutes les entrées dans la base de données électronique, une première étape de sélection a été réalisée selon la pertinence des titres d'articles, puis une deuxième étape de sélection a porté sur l'abstract des articles présélectionnés à la première étape. Pour la troisième et dernière étape de sélection, tous les articles présélectionnés à la deuxième étape ont été sélectionnés sur texte intégral.

Après suppression des doublons, la recherche via Pubmed a aboutit à 1152 entrées. Après analyse du titre et de l'abstract des articles, 225 articles semblaient correspondre au

sujet de la recherche et le texte intégral de ces articles a été analysé en détail. Finalement, après sélection sur texte intégral et consensus des relecteurs, 212 articles ont été inclus dans la revue de littérature.

Les 212 articles analysaient le reporting d'études de 37 spécialités médicales, paramédicales ou de médecines alternatives différentes. 34 articles (16% des publications) analysaient la qualité du reporting des études, sans se limiter à un domaine médical, paramédical ou de médecine alternative précis, représentant ainsi le sous-groupe d'articles le plus représenté dans cette revue de littérature [96-129].

Les disciplines les plus analysées en termes de qualité de reporting, étaient : L'oncologie (22 articles, 10.4% des publications [130-151]), la dentisterie/odontologie (17 articles, 8% des publications [152-168]), l'acupuncture (14 articles, 6.6% des publications [169-182]) et la pédiatrie (10 articles, 4.7% des publications [183-192]).

3-2-1 Evaluation globale de la qualité du reporting des études

34 articles parmi les 212 s'intéressaient particulièrement à la qualité globale du reporting des études, en s'aidant des différentes recommandations de reporting : 17 avaient analysé spécifiquement les essais cliniques randomisés [99,100,103,104,106,107,110,115,117,118,123-129](déclaration CONSORT et extensions), 5 avaient étudié les revues systématiques de littérature et les méta-analyse [102,109,110,113,121] (déclaration PRISMA et extensions), 4 se sont focalisées sur les études observationnelles [96,108,119,122] (déclaration STROBE et extensions), 4 avaient analysé exclusivement les abstracts d'essais cliniques randomisés [98,101,111,120] (extension pour abstracts de la déclaration CONSORT), 2 avaient évalué la qualité du reporting de plusieurs types d'études différentes [97,116] (entre 5 et 8 déclarations de reporting utilisées selon les études) , une s'est concentrée sur les études d'association génétique [105] (déclaration STREGA) et une avait analysé les essais de précision diagnostique (déclaration STARD) [114]. Il faut rappeler que l'objectif premier de l'ensemble de ces recommandations était d'aider les auteurs et les éditeurs pour la rédaction et la publication de leurs manuscrits et ne devait pas initialement servir d'outil d'évaluation qualitative.

De nombreuses études constataient que la qualité du reporting des essais cliniques randomisés restait sous optimale : Martin J. et al ont réalisé une revue systématique de littérature portant sur la qualité du reporting du calcul de taille d'échantillon, dans des essais cliniques en grappes à plusieurs étapes [106]. Les auteurs ont utilisé pour l'analyse, l'extension de la déclaration CONSORT dédiée aux essais en grappe (publiée en 2012, accessible via l'EQUATOR network [75]). Au total, 32 essais achevés et 28 protocoles indépendants ont été inclus. 45 essais (70% des cas) avaient correctement rapporté le calcul du nombre de sujets dans la section matériels et méthodes. Cependant, aucun essai n'avait rapporté la totalité des 9 items requis dans CONSORT pour calcul de ce nombre de sujets à inclure. Seulement 26.7% des essais avaient clairement rapporté si le design de l'essai était transversal ou de cohorte. En conclusion, les auteurs avaient constaté une qualité sous-optimale des items de reporting relatifs au calcul du nombre de sujet dans ce type d'essais.

Ces résultats sont cohérents avec ceux publiés par Rutteford C et al [110], où 166 essais cliniques en grappe (55% des cas) avaient rapporté le calcul de taille d'échantillon.

Dans l'article de Clark L. et al [117], la qualité du reporting méthodologique des essais cliniques randomisés publiés en Janvier 2011 a été évaluée, notamment en ce qui concerne le double aveugle et le processus de randomisation. Sur les 85 essais inclus, 63 (74% des cas) avaient explicitement adhéré à la déclaration CONSORT. L'item portant sur les méthodes de génération de séquences était rapporté dans 74% des essais. En revanche, l'item portant sur la description des procédures de dissimulation d'allocation à un bras de traitement (expérimental ou traitement de référence) n'était correctement rapporté que dans 41% des essais. Seulement 38% des essais avaient rapporté clairement des détails portant sur les personnes en charge d'inclure les participants et de générer la séquence de randomisation.

Ce même défaut de reporting a été également observé dans l'article de Turner L. et al [118], où l'item portant sur la dissimulation de l'allocation à un bras de traitement donné n'était correctement rapporté que dans 22% des essais cliniques randomisés, publiés dans des journaux n'adhérant pas à la déclaration CONSORT (329/1520 essais). Cet item était rapporté plus souvent dans le groupe des essais publiés dans les journaux adhérant au CONSORT : 45% des essais (393/876 essais, RR=1.81, IC99[1.25-2.61]). Dans cette revue systématique, d'autres items de reporting étaient plus souvent cités dans les journaux adhérents versus non adhérents au CONSORT : Rationnel scientifique et contexte (RR 1.07, IC99 [1.01-1.14]), calcul de la taille d'échantillon (RR 1.61, IC99[1.13-2.29]), méthode de génération de séquence (RR 1.59, IC99[1.38-1.84]). En conclusion, les auteurs avançaient l'hypothèse que l'adhésion des journaux médicaux à la déclaration CONSORT pourrait avoir un effet bénéfique sur la qualité du reporting des essais publiés.

Le reporting des essais cliniques randomisés présente donc des lacunes, essentiellement pour les items d'ordre méthodologique. Les journaux n'adhèrent pas tous à la déclaration CONSORT. La qualité globale du reporting des essais randomisés s'est améliorée depuis l'an 2000, mais reste à un niveau insuffisant [128].

La qualité de reporting des revues systématiques, via la déclaration PRISMA, a été évaluée dans l'article de Page MJ. Et al en 2017[102]. Cette étude a porté sur 100 méta-revues, qui avaient investigué certains items de la déclaration PRISMA ou d'une de ses extensions. 57 méta-revues (57% des cas) avaient analysé l'adhésion de 6487 revues systématiques à PRISMA. Les résultats de ces études suggéraient que le reporting de nombreux items de la déclaration PRISMA était sous-optimal, même pour les 2382 revues systématiques publiées après 2009 (date de publication de PRISMA) où 9 items étaient rapportés dans moins de 67% des cas. En conclusion, pour améliorer la transparence des revues systématiques, les auteurs proposent une mise à jour de la déclaration PRISMA, ainsi qu'un ensemble de stratégies pour aider les journaux à adhérer et implémenter la déclaration mise à jour. Willis BH et al avaient étudié la qualité du reporting de méta-analyses de tests diagnostiques publiées, en utilisant le PRISMA [121]. 236 méta-analyses ont été prises en compte pour l'analyse. La qualité du reporting de ces études étaient globalement médiocre : Aucune n'avait respecté l'ensemble des 27 items de la liste PRISMA, 2 (1% des cas) avaient

rapporté le protocole de l'étude, 59 (25% des cas) avaient rapporté la stratégie de recherche utilisée, 76 (32% des cas) avaient rapporté les résultats d'évaluation du risque de biais et 82 (35% des cas) avaient rapporté l'abstract sous la forme d'un résumé structuré. Toutefois, il y avait une amélioration du reporting de certains items entre les méta-analyses publiées entre 2001 et 2004 (63 études) et celles publiées entre 2005 et 2008 (156 études) : Critères d'éligibilité (RR 1.13 IC95[1-1.27]), risque de biais entre les études(méthodes) (RR 1.81 IC95[1.34-2.44]), résultats sur la sélection des études (RR 1.48 IC95[1.05-2.09]), résultats des études individuelles (RR 1.37 IC95[1.09-1.72]), risque de biais entre les études(résultats) (RR 1.65 IC95[1.20-2.25]).

Certains items de la déclaration PRISMA restent donc peu rapportés dans les revues systématiques et les méta-analyses, ce qui nuit à la transparence et aussi à la reproductibilité de telles études.

Concernant les études observationnelles, Pouwels KB. et al [108] ont publié une revue de littérature, portant sur la qualité du reporting des éventuels facteurs de confusion, dans 394 études de cohorte ou cas-témoins publiées entre janvier 2010 et décembre 2012. Six items de reporting portant sur les facteurs de confusion ont été notés pour chaque article. Le nombre médian d'items correctement rapportés dans les études était assez similaire, qu'elles soient publiées avant (174 études) ou après (220 études) la publication de la déclaration STROBE (médiane = 4 ; Q1-Q3 [3-5] versus médiane= 4 ; Q1-Q3 [3.75-5]). En guise de conclusion, les auteurs affirmaient que le reporting global des études observationnelles s'était amélioré au cours des dernières années, mais restait malgré tout imparfait. En 2018, Sharp MK. Et al avaient publié dans le Journal of clinical epidemiology une étude dont l'objectif principal était d'estimer la prévalence et la typologie de l'adhésion à la déclaration STROBE des journaux médicaux [96]. Sur un total de 257 journaux différents publiant des études observationnelles, 139 (54% des cas) ne mentionnaient nulle part la déclaration STROBE dans les instructions aux auteurs, tandis que seulement 12 journaux (5% des cas) exigeaient que les items de STROBE soient respectés au moment de la soumission. L'adhésion à d'un journal à la déclaration STROBE était associée à son adhésion au CONSORT et au PRISMA.

En conclusion, l'adhésion des journaux au STROBE reste faible, l'information destinée aux auteurs reste souvent vague et ambiguë. L'amélioration de la qualité du reporting des études observationnelles semble donc passer tout d'abord par un effort de la part des éditeurs pour prendre l'adhésion au STROBE dans leur évaluation des articles soumis, et de prodiguer un message clair aux auteurs dans ce sens.

Plusieurs études se sont intéressées plus particulièrement à la qualité du reporting des abstracts d'essais cliniques randomisés. Chhapola V. et al ont réalisé une revue systématique de littérature, portant sur la comparaison du reporting de 5184 abstracts d'essais cliniques randomisés publiés avant ou après la parution de l'extension CONSORT dédiée aux abstracts [101]. Le reporting post CONSORT était évalué comme sous optimal pour 10 items (titre, participants, critère de jugement principal, détails sur l'aveugle, statut industriel ou académique de l'essai, nombre de patients analysés, résultats sur le critère principal, précision et taille de l'effet, toxicités, numéro d'enregistrement de l'essai). En conclusion, les auteurs

considéraient que la modification du reporting des abstracts d'essais clinique randomisés était loin d'être satisfaisante depuis la publication de l'extension CONSORT abstract.

3-2-2 Evaluation de la qualité du reporting des études en dentisterie

Sur les 212 articles, 17 avaient porté sur le sujet de la qualité du reporting des études dans le domaine de la dentisterie et/ou odontologie [152-168]. Parmi ces articles, 11 avaient analysé la qualité du reporting des essais cliniques randomisés [153,155-158,160, 162,165-168], 2 se sont focalisées sur la qualité du reporting des revues systématiques et méta-analyses [152,154], 2 ont étudiés le reporting des études observationnelles [163,164] et les 2 derniers ont porté sur la qualité du reporting des abstracts (essais cliniques randomisés [161], revues systématiques et méta-analyses [159]).

En ce qui concerne les essais cliniques randomisés en dentisterie et/ou odontologie, l'article de Savithra P et al [160] a analysé 114 essais cliniques randomisés, publiés entre 1990 et 2009. Les recommandations CONSORT étaient recherchées pour chaque publication. La conclusion des auteurs était que la qualité de reporting des essais cliniques randomisés était médiocre dans les journaux de dentisterie. De plus, la qualité globale de reporting ne semblait pas améliorée depuis la publication de CONSORT. En 2017, Lucena C et al [156] avaient observé que de nombreux items d'ordre méthodologique étaient peu rapportés dans les études d'essais cliniques en odontologie, tels que la méthode de dissimulation d'allocation du traitement (non rapporté dans 75% des études), le reporting de l'intervalle de confiance associé au résultats (rapporté dans 16% des études), le reporting exact de la valeur de la p-value (non rapporté dans 25% des études). Pour 87% des études analysées, l'information contenue dans la publication a été considérée comme insuffisante afin de déterminer si la méthodologie statistique appliquée semblait appropriée ou non. Alamri HM et al [153] avaient inclus dans leur revue systématique un total de 109 publications d'essais cliniques randomisés d'endodontie, publiés entre 2012 et 2017. Le score moyen de qualité globale du reporting était égal à 65%. Sur la totalité des publications, seulement 56 (51 % des cas) avaient rapporté de manière claire les 5 items relatifs à la méthode de randomisation. 23% des études ne reportaient pas leur protocole dans la publication.

Concernant la qualité du reporting des revues systématiques en dentisterie, Bassani R. et al [152] avaient souligné un défaut de reporting de l'ordre de 40.5% pour l'item risque de biais/méthode d'évaluation de la qualité des études, parmi les 495 articles de revues analysés. Les auteurs avaient identifié 4 items qui étaient plus souvent rapportés dans les revues de littérature utilisant la déclaration PRISMA : L'item description revues systématique ou méta-analyse dans le titre/abstract, l'item éligibilité à la publication des études incluses dans l'analyse, l'item description complète de la stratégie de recherche booléenne et l'item critères de jugement spécifiés dans la partie Méthodes. Dans l'article de Lee DW et al [154], 21 méta analyses de réseau dans le domaine des soins dentaires ont été analysés. 15 (71%) des méta analyses ne décrivaient pas en détail le réseau, comme stipulé sur l'extension PRISMA.

3-2-3 Evaluation de la qualité du reporting des études en acupuncture

Un total de 14 études s'est focalisé sur la qualité du reporting des études réalisées en acupuncture [169-182]. Parmi ces études, 12 ont porté sur la qualité du reporting des essais cliniques randomisés [169-175,177-179,181,182], une sur la qualité du reporting des revues systématiques de littérature ou méta-analyses [176] et une a analysé la qualité du reporting des études observationnelles et séries de cas en acupuncture [180].

Dans l'article de Lu X et al [181], les auteurs ont évalué la qualité du reporting d'essais cliniques randomisés portant sur le sujet du traitement par acupuncture des troubles cognitifs modérés. 14 essais ont été inclus dans l'étude. En se basant sur la déclaration CONSORT, 7 essais (50% des cas) avaient correctement rapporté la méthode utilisée afin de générer la séquence de randomisation aléatoire, et seulement 2 essais (14.3% des cas) avaient rapporté une technique adéquate de dissimulation d'allocation du traitement. Ensuite, des items de reporting plus spécifiques aux essais d'acupuncture ont été analysés, en se basant sur la déclaration STRICTA, qui est une recommandation de reporting conçue spécifiquement pour les essais cliniques d'acupuncture (accessible via le site de l'EQUATOR network [75]). 6 études (42.86% des cas) avaient rapporté clairement la réponse au traitement d'acupuncture et 3 études (21.43% des cas) avaient rapporté précisément le nombre d'aiguilles utilisées pour le traitement. En conclusion, les auteurs avaient considéré la qualité du reporting de ces études comme moyenne à mauvaise, avec une utilisation trop peu répandue des recommandations de reporting CONSORT et STRICTA.

En 2018, Jia P et al [169] ont publié un article se focalisant sur la qualité de reporting des essais cliniques randomisés d'acupuncture dans le traitement de l'arthrose du genou. 318 rapports d'essais ont été étudiés. En utilisant la checklist CONSORT, 10 items ont été considérés comme peu rapportés dans ces études : le(s) éventuel(s) changement(s) de la partie Méthodes après le début de l'essai (présent dans 0.6% des cas), description de tout changement éventuel de critère de jugement au cours de l'étude (présent dans 0% des cas), implémentation d'analyses intermédiaires et de règles d'arrêt éventuel de l'étude (présent dans 1.6% des cas), phrase portant sur l'enregistrement de l'étude (présent dans 4.4% des cas), accessibilité au protocole complet de l'étude (présent dans 4.7% des cas), détails sur l'implémentation de la randomisation (présent dans 4.7% des cas), description de la similitude des interventions (présent dans 3.5% des cas), réalisation ou non d'analyses ancillaires (présent dans 3.8% des cas) et présentation des méthodes utilisées pour les analyses additionnelles (présent dans 4.4% des cas). Concernant les items de la déclaration STRICTA, 4 ont été considérés comme peu rapportés (dans moins de 10% des essais cliniques randomisés) : Description du style d'acupuncture, présentation de l'ampleur de la variation du traitement selon les praticiens, niveau d'expertise du (ou des) praticien(s) ayant participé à l'étude, rationnel du choix d'un groupe contrôle. La conclusion des auteurs était assez similaire à l'article précédent, où la qualité de reporting des études analysées demeurait assez faible et où il était réitéré aux journaux des conseils d'adhérence plus stricte aux déclarations CONSORT et STRICTA.

3-2-4 Evaluation de la qualité du reporting des études en pédiatrie

Dans la présente revue de littérature, il a été identifié 10 publications portant sur la qualité du reporting des études en pédiatrie [183-192]. Parmi ces publications, 6 avaient pour sujet la qualité du reporting des essais cliniques randomisés de pédiatrie [183-186,190,191], 2 avaient porté sur la qualité de reporting des revues systématiques de littérature [189,192], une sur la qualité des études observationnelles [187] et une sur le reporting des abstracts d'essais de pédiatrie [188].

Dans l'article de Devos F et al [184], la qualité du reporting des essais cliniques randomisés de pédiatrie publiés dans des journaux de soins infirmiers a été comparée à ceux publiés dans des journaux médicaux. Les auteurs ont utilisé la déclaration CONSORT pour évaluer cette qualité de reporting, en utilisant un score qui prenait en compte 28 items de la checklist CONSORT. Un total de 56 articles médicaux et de 66 articles infirmiers ont été inclus dans l'analyse. Le score CONSORT global n'était pas statistiquement différent selon le type de journal : 19.2 [16.2; 22] pour les journaux médicaux et 19.5 [16.1; 21.5] pour les journaux infirmiers, $p = 0.77$. Il y avait une différence statistiquement significative pour seulement 2 items, qui étaient plus souvent rapportés dans les journaux médicaux (item 19 : déclaration de tout effet indésirable grave ou effet inattendu $p=0.0006$, item 23 : numéro d'enregistrement de l'étude $p=0.0003$). Les deux facteurs principaux liés à une meilleure qualité de reporting étaient l'adhérence du journal à la déclaration CONSORT et une taille d'échantillon importante dans l'étude. Les auteurs avaient considéré que le reporting des items de CONSORT dans ces publications était mauvais, et ce, quel que soit le type de journal (médical ou infirmier) à l'origine de ces publications. En 2010, Brooks RJ et al [190] ont publié un article portant sur la qualité du reporting des essais cliniques randomisés de transplantation rénale en pédiatrie. Les auteurs se sont basés sur une checklist de 22 items de CONSORT afin de réaliser cette évaluation. 27 publications ont été analysées. Le score CONSORT moyen était égal à 67%, avec des lacunes importantes dans les parties méthodes, résultats et discussion des articles. En conclusion, les auteurs ont considéré que des efforts conséquents devaient être réalisés afin d'optimiser la qualité du reporting des publications dans ce domaine. Al faleh K et al [192] ont publié en 2009 un article qui s'était focalisé sur la qualité du reporting des revues systématiques publiées de la Cochrane Neonatal Review Group (CNRG). La qualité du reporting avait été évaluée à l'époque en utilisant la déclaration QUOROM, déclaration de reporting qui servait de référence avant la publication de PRISMA. 61 revues systématiques avaient été analysées. Les principaux points d'amélioration mis en évidence portaient sur le reporting du titre (identification claire de l'étude comme étant une revue systématique de littérature) et de l'abstract des manuscrits (reporting dans l'abstract des principales caractéristiques des études analysées dans la revue, reporting des analyses de sous-groupe). La plupart des revues ne rapportaient pas adéquatement leur évaluation clinique et statistique d'hétérogénéité inter étude (21% de bon reporting), tout comme leur rationnel justifiant leurs analyses de sous-groupes prévues a priori (11% de bon reporting). En conclusion, les revues systématiques publiées dans le CNRG ont une bonne qualité de reporting, avec des défauts mineurs.

3-2-5 Evaluation de la qualité du reporting des études en oncologie

En 2018, Tardy MP et al [132] ont publié un article dont le sujet était l'évaluation de la qualité du reporting des publications d'essais cliniques randomisés réalisés chez les patients adultes atteints d'un gliome de haut grade. L'ensemble des publications Pubmed entre Janvier 1990 et Décembre 2016 ont été prises en compte pour l'analyse. La mesure de qualité du reporting était basée sur la déclaration CONSORT, et les auteurs ont élaboré un score spécifique appelé CONSORT score (CS), où chaque item de reporting évalué était coté « 1 » si présent et « 0 » si non présent dans le manuscrit. 96 publications ont été étudiées. Le score CS médian était égal à 19.5 (min-max 3-30). Certains items étaient rapportés dans moins de 25% des publications, tels que la méthode de randomisation utilisée ou des détails sur le maintien du simple ou double aveugle au cours de l'étude. Ce score CS était généralement plus élevé dans les publications récentes : Score CS médian = 13 entre 1990 et 1996 (avant la publication de CONSORT), score CS médian=17 entre 1996 et 2004 et score CS médian=22 après 2005. En analyse multivariée, les facteurs liés à un score CS plus élevé étaient : La publication de l'essai clinique dans un journal dont l'impact factor est supérieur à 10 ($p<0.001$), une publication de l'étude après 2010 ($p=0.001$), une définition claire du critère de jugement principal de l'essai ($p<0.001$), une inclusion de plus de 200 patients dans l'essai randomisé ($p=0.004$). En conclusion, les auteurs considèrent que la qualité globale du reporting des publications d'essais randomisés chez les patients atteints de gliome de haut grade s'est améliorée au fil du temps, mais qu'il persiste des lacunes concernant certains items de reporting tels que la description de la méthode de randomisation et la description sur le maintien de l'aveugle au cours de l'étude.

Chen YP et al [134] ont analysé la qualité du reporting d'articles d'essais cliniques randomisés de radio-chimiothérapie chez des patients atteints de cancer du nasopharynx. Un score de reporting, intitulé Overall Quality Score (OQS), sur 28 points et basé sur la déclaration CONSORT, a été élaboré avant l'analyse. 24 articles ont été retenus dans cette étude. Le score OQS médian était égal à 15.5 (min-max :10-24). La moitié des items compris dans le score OQS étaient mal rapportée dans 40% des publications. Après analyse multivariée, les facteurs liés à un meilleur score OSQ étaient une publication de l'article après 2010 et un impact factor élevé du journal dans lequel l'essai clinique était publié. La qualité de reporting des publications d'essais clinique réalisés chez les patients porteurs d'un cancer du nasopharynx était au final considérée par les auteurs comme non satisfaisante.

Concernant la qualité du reporting des abstracts des essais de phase III en oncologie, Sivendran S et al [138] ont réalisé une revue systématique s'intéressant à ce sujet. Les auteurs ont intégré dans leur analyse des abstracts d'essais cliniques de phase III réalisés chez des patients atteints d'un cancer solide métastatique et publiés dans pubmed, embase et medline entre janvier 2009 et décembre 2011. Il a ensuite été élaboré un score de qualité de reporting, sur 18 points et basé sur l'extension 2007 de la déclaration CONSORT dédiée au reporting des abstracts. 174 abstracts ont ainsi été étudiés, le score médian de reporting était égal à 9 (min-max 3-17). Il y avait un accès libre à l'article en texte intégral pour 80% des abstracts. Les items de reporting les moins souvent retrouvés étaient : La description du design de l'étude (présent dans 20% des abstracts), la description de la méthode d'allocation des

traitements (présent dans 13% des abstracts), description de l'aveugle (présent dans 24% des abstracts), le statut de l'étude vis-à-vis des inclusions (détaillé dans 22% des abstracts), le nom et numéro d'enregistrement de l'étude (disponible dans 26% des abstracts) et le type de financement (présent dans 18% des abstracts). En conclusion, les auteurs ont constaté une faible adhérence aux recommandations CONSORT dédiée aux abstracts. L'article de Guo et al [144] confirme ce reporting sous optimal des abstracts des études en oncologie : Sur un total de 227 abstracts d'études infirmières en oncologie, plus de 80% décrivaient seulement 6 des 17 items de reporting de la checklist CONSORT 2007 dédiée aux abstracts. Des items méthodologiques importants, tels que la description de la randomisation, la description de l'aveugle et la précision sur l'analyse en intention de traiter, étaient correctement rapportés dans moins de 30% des abstracts.

Un seul article s'est focalisé sur la qualité du reporting des publications de méta-analyses de réseau en cancérologie [135]. Les auteurs se sont intéressés non seulement à la qualité du reporting de ces articles, en se basant sur la déclaration PRISMA, mais également à la qualité méthodologique, en utilisant la checklist AMSTAR (Assessing the Methodological quality of Systematic reviews). 102 publications ont été identifiées et incluses dans l'analyse. Le score AMSTAR médian était égal à 8 (interquartile 6-8.25) et considéré comme satisfaisant par les auteurs. Concernant les éléments de reporting, 13 publications de méta-analyses (12.7% des articles) n'avaient pas rapporté d'information sur la stratégie de recherche de la littérature. Sur les 61 méta-analyses utilisant une méthodologie Bayésienne, 60.6% des publications ne rapportaient pas d'évaluation de convergence. La plupart des articles n'avaient pas rapporté d'évaluation de similarité (86.89% des cas). Seulement 4.65% des publications avaient rapporté clairement les détails sur le traitement des essais multi-groupes.

Une publication a porté sur la qualité du reporting d'études médico-économiques [136] : Dans cet article, les auteurs ont réalisé une revue systématique des études d'évaluation économique, portant sur les nouvelles thérapies disponibles dans le traitement du myélome multiple. L'objectif principal a consisté à évaluer la qualité de reporting de ces études médico-économiques, en utilisant la déclaration CHEERS (Consolidated Health Economic Evaluation Reporting Standards, accessible via le site de l'EQUATOR network [75]). Au total, 8 études correspondaient aux critères d'inclusion et ont été analysées. La plupart des études étaient des analyses coût-efficacité combinées à des analyses coût-utilité (n = 6) avec perspective de payeur public (n = 4). Elles ont été réalisées essentiellement en Europe (n = 6) sur des patients atteints de myélome multiple réfractaire ou en rechute (n = 5). Dans l'ensemble, l'intervention a été plus efficace et plus coûteuse que la stratégie alternative (moyenne de 54 630\$ par année de vie; 68 261\$ par année de vie ajustée en fonction de la qualité - QALY). Le score total CHEERS moyen était égal à 14.6/24 (Ecart-Type = 2,6). Les items les moins souvent rapportés étaient : Les mesures de précision pour tous les paramètres du modèle (présent dans 25% des études), l'évaluation de l'hétérogénéité des résultats par analyse de sous-groupe (présent dans 0% des études), la description du rôle du bailleur de fonds dans l'identification, la conception, la réalisation et le compte rendu de l'analyse (présent dans 12.5% des études). En conclusion, Aguiar PM et al ont considéré la qualité de reporting de ces études médico-

économiques comme médiocre et que des efforts doivent être réalisés afin d'assurer une plus grande transparence.

En conclusion, quel que soit la discipline concernée, les recommandations de reporting déjà publiées et accessibles pour les auteurs et pour les comités de relecture de journaux médicaux (CONSORT, STROBE, PRISMA, CHEERS, etc) sont suivies de façon inconstante. Ceci impacte sur la transparence des informations contenues dans les différents types de publications. Certains éléments méthodologiques importants ne sont ainsi pas clairement rapportés dans les articles, ce qui peut nuire à une éventuelle reproductibilité de l'étude ou à la réalisation de méta-analyses en cas d'information manquante. Plusieurs études ont également mis en évidence l'absence d'amélioration du reporting au cours du temps [102,160], ce qui reflète la persistance d'une certaine hétérogénéité dans l'adhésion aux recommandations de reporting selon les journaux médicaux [96,118]. L'oncologie demeure la spécialité où le reporting des articles d'études a été le plus analysé (22 articles, [130-151]), avec un intérêt plus important pour les publications d'essais cliniques randomisés.

Avec l'avènement récent de traitements innovants, telles que les thérapies moléculaires ciblées, l'immunothérapie, les nouvelles techniques de radiothérapie, le nombre d'essais cliniques en oncologie a augmenté de façon importante depuis les années 2000 [10]. Le reporting des publications d'études en oncologie représente ainsi un sujet de recherche important, afin d'optimiser la conception et la méthodologie des essais, dans un souci de transparence constant. Le travail de thèse s'est donc concentré sur le reporting des études en oncologie, discipline où les progrès thérapeutiques sont constants et où la recherche clinique est très active.

II- Hypothèses de recherche et objectifs de la thèse

Lors de la réalisation du travail d'introduction de thèse, dans la partie relative aux spécificités méthodologiques des essais en oncologie, il a été précisé l'importance de respecter une méthodologie rigoureuse, adaptée à l'hypothèse de recherche et à l'objectif de la recherche, lors de chaque étape du processus de développement d'un nouveau médicament ou d'une nouvelle stratégie thérapeutique. En effet, la mise en œuvre d'une méthode de recherche inadaptée, peut impacter la validité des résultats obtenus et, à termes, une potentielle autorisation de mise sur le marché. Chacune des différentes phases de développement : les études observationnelles d'état des lieux, l'essai de phase II, la phase III, les études observationnelles de phase IV, constituent l'ensemble des étapes cruciales pour le développement d'un nouveau médicament. Un état des lieux qui conclue à une fausse interprétation peut orienter une recherche de développement de médicament dans une mauvaise direction. De même pour les phase II, les molécules n'ayant pas démontré une efficacité minimale au décours de cette phase, ne pourront pas poursuivre leur développement dans des essais comparatifs de phase III de plus grande échelle. Des essais de phase III concluant de façon erronée sur l'efficacité ou la non-efficacité d'un médicament, pour des raisons purement méthodologiques, impacteront de façon négative le devenir des patients

concernés. Malgré l'importance de chacune de ces étapes, il n'a pas été objectivé dans le travail d'introduction de thèse, de recommandations de reporting spécifique aux articles d'essais de phase II. De plus, aucune revue de littérature ne semble avoir été réalisée pour synthétiser l'ensemble des données connues sur le reporting des publications d'études observationnelles, d'essais de phase II ou d'essais de phase III.

L'ensemble de ces données ont incité à formuler les hypothèses de recherche suivantes :

1. Pour les schémas expérimentaux où des recommandations internationales de reporting existent (STROBE pour les études observationnelles, CONSORT pour les essais contrôlés randomisés), la littérature s'intéressant à la qualité du reporting de ces études devrait être riche dans le domaine de l'oncologie. Il est toutefois probable que pour ces 2 types de design, il reste des éléments structurels ou méthodologiques— spécifiques au domaine de l'oncologie - qui soient moins étudiés par cette littérature.
2. Pour les schémas expérimentaux développés particulièrement en oncologie tels que les phases II, où il n'existe pas de recommandations internationales de reporting, il est attendu que la littérature s'intéressant à la qualité du reporting de ces études soit pauvre et/ou identifie une grande variabilité dans le reporting et sa qualité.
3. Pour les études interventionnelles de type essai contrôlé randomisé ayant des critères de jugements non communs et pour les études observationnelles, il est attendue une association entre des éléments intrinsèques des études, et la qualité du reporting
4. L'absence de recommandations internationales de reporting des phases II a probablement pour conséquence une variation des pratiques en termes de reporting des phases II qui peut probablement s'expliquer par des caractéristiques générales ou d'ordre méthodologique des essais de phase II

Afin de répondre à l'ensemble de ces hypothèses, quatre revues de littérature ont été réalisées.

Le premier article du travail de thèse a consisté à réaliser une revue de l'ensemble des revues de littérature portant sur le reporting des études observationnelles et interventionnelles de type essai contrôlé randomisé en oncologie, censées répondre aux exigences des recommandations CONSORT et STROBE. Ceci afin d'objectiver d'éventuel(s) type(s) d'étude(s) moins étudiée(s) en termes de reporting. Le deuxième objectif était d'apprécier le reporting propre à chaque revue de littérature incluse, en utilisant un score de reporting issu de PRISMA.

L'objectif du deuxième article du travail de thèse a été de synthétiser l'ensemble des données de la littérature disponible sur le reporting des articles d'essais de phase II en oncologie et la qualité de ce reporting. La revue de littérature réalisée a eu également pour objectif de décrire les outils disponibles utilisés dans un but d'optimisation du reporting des essais de phase II en oncologie.

Le troisième article du travail de thèse permet de prolonger et de répondre aux questions soulevées par l'article 1 vis-à-vis des études observationnelles et interventionnelles de type essai contrôlé randomisé : il s'est intéressé à décrire la qualité du reporting du critère

d'évaluation principal dans ces études. La revue de littérature a porté aussi bien sur les études interventionnelles (reporting du critère de jugement principal) que sur les études observationnelles (reporting des variables), ces dernières ayant été peu analysées dans d'autres revues de littératures. Le deuxième objectif de ce travail a consisté à analyser le reporting du test statistique utilisé pour l'analyse du critère d'évaluation principal, aussi bien dans la partie Méthodes que dans la partie Résultats du manuscrit de chaque étude sélectionnée.

Le quatrième article du travail de thèse permet de prolonger et de répondre aux questions soulevées par l'article 2. Il avait pour objectif de décrire les différents éléments de reporting des publications d'essais de phase II, en recherchant notamment les éléments les moins décrits. Il a également analysé si certaines caractéristiques des études, générales ou de méthodologie, (année de publication, site tumoral primitif, type de traitement) avaient un lien avec la qualité du reporting.

III- Les articles originaux évaluant le reporting en oncologie

1. Quality of reporting in oncology studies : A systematic analysis of literature reviews and prospects

1-1 Introduction

La première revue narrative de littérature réalisée avait objectivé que l'oncologie demeurait la discipline la plus analysée en termes de reporting des publications d'essais. Toutefois, aucune étude n'avait été conçue dans le but de synthétiser l'ensemble de ces données disponibles, notamment pour mettre en relief les éventuelles améliorations à apporter et éléments de recherche pour optimiser le reporting des études en oncologie. Le premier article original avait donc été réalisé pour répondre à ces questions, sous la forme d'une revue systématique de l'ensemble des revues de littérature, dont le sujet était la qualité du reporting des publications d'essais contrôlés randomisés et d'études observationnelles en oncologie. L'objectif principal consistait à réaliser une synthèse générale sur la qualité du reporting des études interventionnelles contrôlées randomisées ou observationnelles, études qui possèdent déjà des recommandations internationales de reporting, à savoir le CONSORT et le STROBE.

1-2 Méthodes

Des recherches ont été réalisées via Pubmed et la Cochrane Library, afin d'identifier l'ensemble des revues de littérature d'intérêt. Afin d'être éligibles pour l'analyse, les revues de littératures devaient respecter les critères d'inclusion suivants : Etudes de type revue de littérature, dont le sujet portait sur les essais cliniques interventionnels contrôlés randomisés (essais de phase II et/ou essais de phase III) et/ou les études observationnelles réalisées en oncologie. Les revues devaient être publiées entre Août 1996 (date de publication de la déclaration CONSORT) et Octobre 2014, le critère de jugement principal devait être l'évaluation de la qualité du reporting des études analysées, la langue de publication des revues devait être l'anglais. La sélection des revues de la littérature dans cette étude a d'abord été réalisée de façon indépendante par deux auteurs (Romain Rivoirard et Mathieu Oriol),

d'abord sur titre (première étape de sélection) puis sur l'abstract des revues présélectionnées à la première étape (deuxième étape de sélection). Lors de la troisième et dernière étape, un troisième auteur (Aurélié Bourmaud) a procédé à la sélection finale sur le texte intégral des revues sélectionnées préalablement.

Après la phase de sélection des revues de littérature, un auteur (Romain Rivoirard) a procédé à l'extraction des données, supervisé par un deuxième auteur (Aurélié Bourmaud). Les données suivantes ont été recueillies pour chaque revue de la littérature sélectionnée : Nombre d'études analysées dans la revue, types d'études analysées (essais interventionnels et/ou études observationnelles), année de publication, nom du journal dans lequel la revue a été publiée, type(s) de cancer(s) étudié(s), scores et outils utilisés afin d'évaluer la qualité du reporting, critère de jugement principal pour l'évaluation de la qualité du reporting, principaux résultats de la revue de littérature .

Les auteurs ont également procédé à une évaluation de la qualité de reporting intrinsèque des revues de littératures incluses, en utilisant un score adapté de la déclaration PRISMA : A partir des 27 items de la déclaration PRISMA, six critères principaux, composés de 19 items ont été définis : Méthodologie de sélection des études, détails sur l'extraction des données, méthodes d'analyse, évaluation du risque de biais, description des résultats et analyse objective des limitations de la revue de littérature réalisée. Pour chaque item, un score de qualité de reporting était attribué, selon le barème suivant : 0 = Non rapporté dans l'article, 1= Rapporté de façon incomplète, 2= Reporting clair et complet. Lors de la deuxième étape, un score de qualité globale de reporting (Overall Quality Score for litterature Reviews ou OQSR) a été calculé pour chaque revue de littérature, en additionnant les score des six critères décrits ci-dessus. Le score OQSR avait une valeur minimale et maximale respectivement égale à 0 et 12. Le calcul du score OQSR a été réalisé de façon indépendante par deux auteurs (Romain Rivoirard et Aurélié Bourmaud), en cas de divergence, un troisième auteur était sollicité pour le calcul final du score OQSR (Nicolas Magné). Pour l'analyse finale du score OQSR, l'ensemble des revues de la littérature ont été scindées en deux groupes selon leur année de publication : De 1996 à 2009 ou de 2010 à 2014. Le premier groupe rassemblait ainsi l'ensemble des revues publiées avant la publication de PRISMA, tandis que le second groupe correspondait aux revues publiées après publication de ces recommandations de reporting. Le score OQSR moyen et médian de ces deux groupes ont été calculés et comparés en utilisant le test non paramétrique de Wilcoxon. Le score OQSR moyen était accompagné de son écart-type, tandis que le score OQSR médian était accompagné de l'interquartile correspondant. La valeur de significativité du petit p était fixé à 0.05. Les analyses statistiques avaient été réalisées à l'aide du logiciel R, version 3.1.2. (<http://www.R-project.org/>).

1-3 Résultats

58 articles de revues de littératures ont été inclus. 52 revues (90% des articles) avaient analysé exclusivement la qualité de reporting des essais interventionnels tandis que 1 revue (2% des articles) s'était intéressée exclusivement à la qualité de reporting des études observationnelles. 17 revues (29% des articles) avaient inclus plus de 200 études à analyser. Sur les 58 revues de littérature, 22 (38% des articles) avaient utilisé la déclaration CONSORT

afin d'évaluer la qualité du reporting des études. Le critère de jugement principal le plus souvent retrouvé était la qualité globale de reporting (20 revues, soit 34% des articles), suivi du reporting de données portant sur la qualité de vie des patients inclus dans les essais cliniques randomisés (7 revues, soit 12% des articles) et la qualité de reporting du critère de jugement principal dans les essais cliniques randomisés (6 revues, soit 10% des articles).

Parmi les 20 revues de littérature ayant évalué la qualité globale de reporting des essais en oncologie, 8 (40% des articles) avaient souligné un reporting insuffisant d'items relatifs à la méthodologie des essais cliniques randomisés : Certains items étaient ainsi peu rapportés, comme la description de la randomisation, les détails sur l'implémentation et le maintien de l'aveugle, la description de la méthode utilisée pour dissimuler l'attribution de traitement, la précision si l'analyse était réalisée en intention de traiter. 3 revues de littérature ont évalué la qualité du reporting des abstracts d'études réalisées dans le domaine de l'oncologie, et l'ensemble de ces revues ont jugé la qualité de reporting des abstracts comme insuffisant. Concernant les revues de littérature ayant pour sujet la qualité de reporting du critère de jugement principal dans les essais cliniques randomisés, la moitié soulignait un manque de précision dans la définition initiale de ce critère de jugement principal.

Le score OQSR médian de l'ensemble des 58 articles de revue de littérature était égal à 10 : OQSR_{median} = 10 (Q1-Q3 : 8-11). 5 publications avaient le score OQSR maximal (OQSR=12/12, soit 9% des publications). Pour le premier groupe d'articles de revues publié entre 1996 et 2009 (n=29 revues), le score OQSR moyen était égal à 8 : OQSR_{moyen} = 8 (Ecart-type :2). Ce score OQSR moyen était significativement plus élevé pour les revues publiées entre 2010 et 2014 : OQSR_{moyen} = 10 (Ecart-type :2) avec p= 0.0053 (test de Wilcoxon).

1-4 Discussion

Une proportion importante des revues de littérature incluses avait analysé la qualité globale du reporting des articles d'essais cliniques randomisés en oncologie (20 revues, soit 34% des articles). Ces revues avaient mis en évidence, que certains items méthodologiques essentiels tels que la description de la randomisation, les détails sur la mise en place et le maintien de l'aveugle tout au long de l'étude, sont peu rapportés dans les publications, ceci posant un problème global de transparence et de reproductibilité de ces essais. Le deuxième constat était que la qualité du reporting des études observationnelles avait été peu analysée dans les revues de littérature : Seulement une revue (2% des articles) s'était focalisée exclusivement sur ce sujet. Enfin, aucune revue de littérature n'avait abordé exclusivement le sujet de la qualité du reporting de la partie « analyses statistiques » des études en oncologie, qu'elles soient interventionnelles ou observationnelles. La validité et la robustesse des résultats obtenus dans les études interventionnelles ou observationnelles dépendent en partie du type d'analyses statistiques réalisées, par conséquent il s'agit d'items de reporting importants à vérifier dans les publications. Ce manque de données sur ce sujet dans la littérature a ainsi servi de base de travail pour les deuxième et troisième articles originaux.

La qualité intrinsèque de reporting des 58 revues de littérature a été considérée comme satisfaisante : Le score OQSR médian était égal à 10 et 5 revues avaient un score OQSR

maximal. Il a également été observé une tendance vers l'amélioration de la qualité du reporting des revues systématiques de littérature depuis la publication de la déclaration PRISMA en 2009 : OQSR_{moyen (1996-2009)} = 8 (Ecart-type :2) contre OQSR_{moyen (2010-2014)} = 10 (Ecart-type :2), avec $p= 0.0053$. Ceci est probablement dû à une adhérence croissante des éditeurs de journaux médicaux à la déclaration PRISMA.

En conclusion, cette revue de revues de littérature a objectivé un reporting insuffisant de certains items méthodologiques importants dans les publications d'essais cliniques randomisés en oncologie. La qualité de reporting des publications d'études observationnelles en cancérologie a été peu évaluée. L'évaluation de la qualité du reporting de la partie « analyses statistiques » des études en oncologie a été peu étudiée jusqu'à présent

1-5 Article original

Critical Reviews in Oncology/Hematology 112 (2017) 179–189



Oncology reviews

Quality of reporting in oncology studies: A systematic analysis of literature reviews and prospects



Romain Rivoirard^{a,*}, Aurélie Bourmaud^{b,c}, Mathieu Oriol^{b,d}, Fabien Tinquaut^b,
Benoîte Méry^a, Julien Langrand-Escure^e, Alexis Vallard^e, Pierre Fournel^a,
Nicolas Magné^{d,e,f}, Franck Chauvin^{b,c,d}

^a Department of Medical Oncology, Lucien Neuwirth Cancer Institute, Saint Priest-En-Jarez, France

^b Hygee Center, Regional Center for Cancer Prevention, CIC 1408, Saint Priest en Jarez, France

^c EMR3738, Ciblage Thérapeutique en Oncologie, Claude Bernard University, Lyon, France

^d Jean Monnet University, Saint-Etienne, France

^e Department of Radiotherapy, Lucien Neuwirth Cancer Institute, Saint Priest-En-Jarez, France

^f EMR3738, Laboratoire de Radiobiologie Cellulaire et Moléculaire, Claude Bernard University, Lyon, France

Contents

1. Introduction	180
2. Methods	180
2.1. Search strategy	180
2.2. Study selection	180
2.3. Data extraction and synthesis	180
3. Results	181
3.1. Literature search	181
3.2. Global characteristics of literature review studies	181
3.3. Detailed characteristics of literature reviews, scores, tools and primary endpoints used to assess the quality of reporting	182
3.4. Main results of literature reviews	182
3.5. Results of OQSR and missing items	182
4. Discussion	187
Contributors	187
Funding/Support	188
Conflict of interest statement	188
Acknowledgment	188
References	188

ARTICLE INFO

Article history:

Received 23 September 2015

Received in revised form 19 January 2017

Accepted 14 February 2017

Keywords:

Randomized controlled trial

Observational study

Review

Data reporting

Oncology

ABSTRACT

The present review gives an overview of systematic reviews published in peer reviewed Journals analysing quality of reporting in oncology studies. PUBMED and Cochrane library were searched to identify systematic reviews assessing quality of reporting for randomized controlled trials (RCTs) and observational studies (OBS). Recommendations and primary endpoints used to assess the quality of reporting were described. Intrinsic quality of reporting was analyzed using an Overall Quality Score for literature Reviews (OQSR). Main evaluation themes were overall quality of reporting (20/58) and reporting of Health-Related Quality Of Life (HRQOL) in RCTs (7/58). Reporting recommendations used were not detailed in 56.9% of reviews. Insufficient reporting for the methodological description (randomization, blinding details, and allocation concealment) and the rationale for using specific measure of HRQOL were highlighted. OQSR was significantly higher for reviews published between 2010 and 2014 (after the PRISMA Publication), as compared to those published between 1996–2009 (median OQSR 10 (10–11) versus median OQSR 9 (6–10) respectively, $p=0.0053$). Intrinsic quality of reporting is satisfactory and has been improved in the last years.

© 2017 Elsevier B.V. All rights reserved.

* Corresponding author at: Department of Medical Oncology, Lucien Neuwirth Cancer Institute, 1088 Avenue Albert Raimond, 42271, Saint Priest-En-Jarez, France.
E-mail address: romain.rivoirard@icloire.fr (R. Rivoirard).

1. Introduction

In the world, cancer is a major health problem, mainly due to the aging of the population. Indeed, in 2012, there were 14.1 million of new cases of cancer in the world and 8.2 million deaths caused by cancer (WHO, 2017; Les cancers en France en, 2013).

Among the different types of studies, the Randomized Clinical Trial (RCT) is the most appropriate to evaluate innovative therapies or different treatment strategies (Guyatt et al., 2008). However, a poorly designed RCT along with bad reporting can have adverse consequences for clinical practice, which is then based on incomplete or inaccurate results (Rennie, 2001). The quality of reporting also has a significant impact on the feasibility and quality of studies such as meta-analyses (Moher et al., 1998). Several recommendations were published to standardize the process of writing, allowing authors to identify the information which must appear in their articles (A proposal for structured reporting of randomized controlled trials, 1994; Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature, 1994). The CONSORT statement (Consolidated Standards Of Reporting Trials) (Schulz et al., 2010) is currently the benchmark recommendation for the reporting of RCTs. The aim consisted of improving the overall quality of reporting of clinical trials (Begg et al., 1996). The original version was updated in 2001 (Moher et al., 2001) and again in 2010 (Schulz et al., 2010).

However, RCTs cannot answer all important questions about a given intervention (Von Elm et al., 2007). Observational studies can investigate many medical questions, such as the search for cause of diseases or research into the benefits and harms of medical interventions (Black, 1996). There are also recommendations on the reporting of observational studies: The STROBE statement (Strengthening The Reporting of Observational Studies in Epidemiology) was published in 2007, to improve the quality of reporting of observational studies (Vandenbroucke et al., 2007).

In addition, so as to verify the reporting of studies, a peer review system exists in almost all medical journals, and systematically in those of high impact factors (International Committee of Medical Journal Editors, 2013). One of the main aims of peer review is to improve the quality and transparency of a publication by checking that the reported research has been carried out correctly and that the results presented have been interpreted appropriately (Rennie, 1992).

Many articles have been published, about the correlation between what is reported in the articles and recommendations of CONSORT and STROBE on the one hand (Ivers et al., 2011; Sorensen et al., 2013), and what has been submitted by peer review and which was published after peer review on the other hand (Hopewell et al., 2014). Some literature reviews have analyzed the overall quality of reporting in oncology studies (Toulmonde et al., 2011), while others focused on more specific criteria, such as reporting adverse events of cancer treatments (Péron et al., 2013) or reporting of primary endpoints in RCTs (You et al., 2012). There is currently no review of these literature reviews.

As a result, the present work was performed to establish a synthesis of the quality of reporting of RCTs and observational studies in oncology. We also searched for possible significant criteria which are not currently described in the CONSORT and STROBE statement. Finally, we assessed the intrinsic quality of reporting for the literature reviews by using the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) 2009 checklist (Moher et al., 2009); the quality of reporting for systematic reviews published between 1996 and 2009 was compared to those published between 2010 and 2014.

2. Methods

2.1. Search strategy

We searched in the Medline database (via PubMed) and The Cochrane Library in order to identify all publications of literature reviews whose subject was the reporting assessment of RCTs and observational studies in oncology, in peer reviewed Medical Journals. The last search was performed in February 2015, using the terms “reporting”, “cancer”, “oncology”, as keywords and “epidemiologic study characteristics as topic” as MeSH Term. We have included studies published in English between August 1996 and October 2014, because the first version of CONSORT statement was published in August 1996 (Begg et al., 1996) and October 2014 is the date of more recent literature review included in the analysis.

2.2. Study selection

To be eligible, studies had to meet all of the following inclusion criteria:

1. Study type: Literature review.
2. Literature reviews about RCTs (phase II/phase III trials) and/or observational studies, in an oncologic context.
3. Studies published between August 1996 and October 2014.
4. Outcome: assessment of reporting quality for analyzed studies.
5. Publication language: English.

We have not included in our analysis the literature reviews about the reporting of phase I studies, pharmacogenomics studies, meta-analyses, descriptive studies, prediction model building studies, test validation studies, recommendations and medico-economic studies. We excluded reviews from Journals having no peer review system. Only literature reviews that assessed the quality of reporting exclusively in the field of oncology were included, all literature reviews evaluating the quality of reporting across all medical specialties or in the field of multiple medical specialties were excluded. We also excluded literature reviews which focused on reporting of studies of specific and predefined treatments. Among all hits in electronic databases, a first step of selection was made concerning the relevant titles of articles. Then, during the second step, a second selection on abstracts was performed, among the pre-selected articles in the first step. These first two steps were realized independently by two authors (RR and MO). For the third and final stage, all pre-selected articles in the second step were selected on full text, independently by one author (AB).

2.3. Data extraction and synthesis

The data on study characteristics and results were extracted in standardized tables that had been compiled beforehand. Data extraction was performed by one author (RR) and supervised by a second author (AB). The number of studies analyzed in the literature review, type of analyzed studies (RCTs or/and observational studies), year of publication, name of the journal which published the study, type(s) of cancer(s) which was (were) studied, scores and tools utilized for the reporting assessment, the main criterion for judgment of reporting as well as the main result were extracted for each study.

We also assessed the reporting quality of the included systematic reviews, with a score adapted from the PRISMA statement (Moher et al., 2009). From the 27 items of the PRISMA 2009 checklist, six major criteria, composed of 19 items, were created and analyzed: Selection of studies that considered items 6 and 9 of PRISMA 2009, data extraction (items 7, 8, 10, 11), methods of anal-

ysis (items 13, 14, 16), risk of bias (items 12 and 15), description of results (items 17–23) and objective analysis of the limitations of the literature review (item 25). These six criteria were rated as follows: 0 – no reporting in the article; 1 – incomplete reporting; 2 – full reporting. In a second step, an Overall Quality Score for the literature Reviews (OQSR) was calculated for each study by adding the scores of the 6 criteria described above. This overall score ranged from 0 to 12. The median OQSR for all reviews was then calculated, the result was accompanied by interquartile: (interquartile range: quartile1–quartile3). Notation was made by two authors (romain rivoirard and aurélie bourmaud), who were blinded to each other's results. A third author has adjudicated in case of discrepancy (nicolas magné). For the final OQSR analysis, all included literature reviews were separated into two groups according to their year of publication: 1996–2009 or 2010–2014. The first group gathered all reviews published before publication of PRISMA statement while the second group corresponded to reviews published after publication of these recommendations. The mean and median OQSR score for each of these two groups was calculated and compared using the Wilcoxon nonparametric test. Mean OQSR was accompanied with the corresponding Standard Deviation (SD) and median OQSR with the corresponding interquartile. All p-values are two sided, with a significance level of 0.05. Statistical analyses were performed using R software, version 3.1.2 (<http://www.R-project.org/>).

3. Results

3.1. Literature search

After suppression of duplicates, the search in electronic databases resulted in 3928 hits. After title and abstract screening, 55 articles for the first author (RR) and 62 for the third author (MO) seemed potentially relevant and the full-text of these articles was screened in detail. Finally, after selection on the full text and after suppression of duplicates by the second author (AB), 58

articles were included in the literature review (Toulmonde et al., 2011; Péron et al., 2013; You et al., 2012; Arkenau et al., 2011; Bekelman and Yahalom, 2009; Blencowe et al., 2012; Blencowe et al., 2014; Bottomley et al., 2003; Burton and Altman, 2004; Bylicki et al., 2015; Cherny et al., 2009; Claassens et al., 2011; Dirven et al., 2014; Duff et al., 2010; Efficace and Bottomley, 2002; Efficace et al., 2004; Efficace et al., 2006; Efficace et al., 2007; Efficace et al., 2014a; Efficace et al., 2014b; Ghimire et al., 2014; Grellety et al., 2014; Guo et al., 2014; Hopewell and Clarke, 2005; Hui et al., 2012; Kober et al., 2006; Krzyzanowska et al., 2004; Lai et al., 2006; Le Tourneau et al., 2009; Lee and Chi, 2000; Lu et al., 2011; Lu et al., 2013; Martin et al., 2006; Mathoulin-Pelissier et al., 2008; Mhaskar et al., 2012; Mitchell et al., 2009; Papatheasiou and Zintzaras, 2010; Pat et al., 2008; Péron et al., 2012; Perrone et al., 2008; Potter et al., 2011; Riechelmann et al., 2008; Scharf and Colevas, 2006; Seruga et al., 2011; Shitara et al., 2012; Sivendran et al., 2014; Soares et al., 2004; Sorbye et al., 2007; Süt et al., 2008; Swanson and Bailar, 2002; Tanaka et al., 2012; Thezenas et al., 2004; Thoma et al., 2009; Treweek, 2011; Tuech et al., 2005; Vera-Badillo et al., 2013; Zikos et al., 2014; Ziogas and Zintzaras, 2009). The flowchart illustrates the selection process (Fig. 1).

3.2. Global characteristics of literature review studies

Among the 58 literature reviews, 52 (90%) analyzed the quality of reporting of RCTs. 1 review (2%) analyzed the reporting of observational studies (OBS). Almost a third of the literature reviews (29%) included more than 200 studies. Half of the reviews were already available before or in the year of publication of the PRISMA statement (Moher et al., 2009). 62% of literature reviews were published in five journals (*Journal of Clinical Oncology*, *Annals of Oncology*, *Journal of the National Cancer Institute*, *European Journal of Cancer* and *Journal of Clinical Epidemiology*). Data are resumed in Table 1.

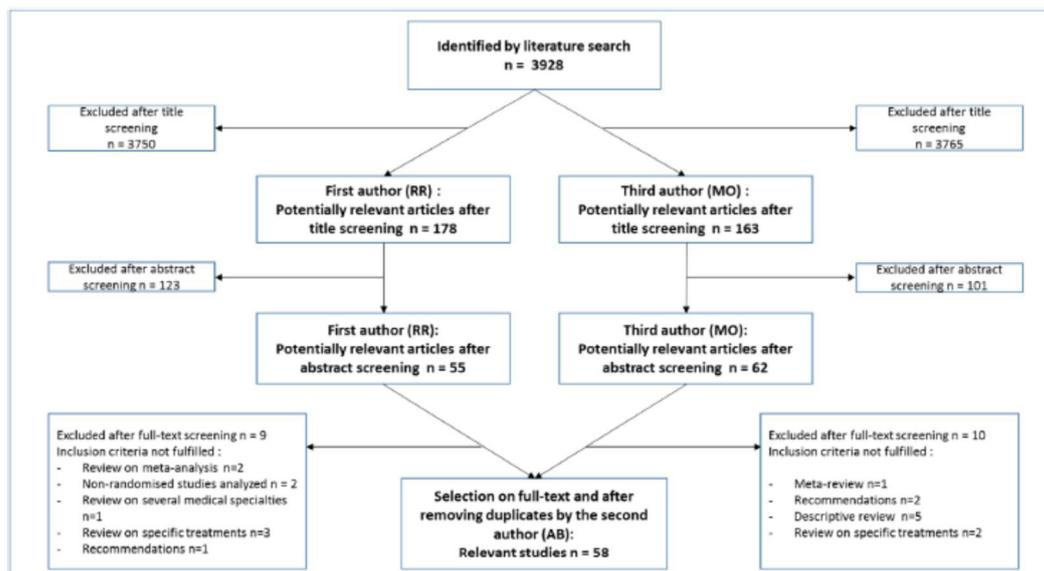


Fig. 1. Identification of literature review studies from Pubmed and Cochrane library search.

Table 1
Characteristics of literature review studies (N = 58).

Characteristic	Literature review, n (%)
Type of analyzed studies	
Randomized Clinical Trials (RCTs)	52 (90)
RCTs and Observational studies (OBS)	5 (9)
OBS	1 (2)
Number of analyzed studies	
<100 studies	28 (48)
101–200 studies	13 (22)
>200 studies	17 (29)
Year of publication	
2000–2009	29 (50)
2010–2012	16 (28)
2013–2014	13 (22)
Journal	
Journal of Clinical Oncology	15 (26)
Annals of Oncology	5 (9)
Journal of the National Cancer Institute	5 (9)
European Journal of Cancer	5 (9)
Journal of Clinical Epidemiology	5 (9)
Cancer	3 (5)
Other journals	20 (34)
Cancer type studied in literature reviews	
All types of cancers	31 (53)
Breast	6 (10)
Colon/Rectum	5 (9)
Others	16 (28)

3.3. Detailed characteristics of literature reviews, scores, tools and primary endpoints used to assess the quality of reporting

Of the 58 reviews, 22 (38%) were based on the CONSORT checklist to assess the quality of reporting. Several reviews then developed a specific evaluation score, usually named Overall Quality Score (OQS: 7/58 reviews, 12.1% of publications). 20/58 literature reviews (34%) had the overall quality of reporting as primary endpoint (Toulmonde et al., 2011; Ghimire et al., 2014; Grellety et al., 2014; Guo et al., 2014; Hopewell and Clarke, 2005; Hui et al., 2012; Kober et al., 2006; Krzyzanowska et al., 2004; Lai et al., 2006; Lu et al., 2011; Lu et al., 2013; Papathanasiou and Zintzaras, 2010; Pat et al., 2008; Péron et al., 2012; Soares et al., 2004; Süt et al., 2008; Tanaka et al., 2012; Thoma et al., 2009; Tuech et al., 2005; Ziogas and Zintzaras, 2009). Other reviews mainly analyzed: Reporting of Health-Related Quality Of Life (HRQOL) in RCTs (7/58 reviews, 12% of publications) (Bottomley et al., 2003; Claassens et al., 2011; Efficace and Bottomley, 2002; Efficace et al., 2004; Efficace et al., 2007; Lee and Chi, 2000; Zikos et al., 2014), the quality of reporting of primary endpoints in RCTs (6/58 reviews, 10% of publications) (You et al., 2012; Arkenau et al., 2011; Le Tourneau et al., 2009; Mathoulin-Pelissier et al., 2008; Riechelmann et al., 2008; Vera-Badillo et al., 2013) and the quality of reporting of adverse events in phase II and phase III studies (5/58 reviews, 9% of publications) (Péron et al., 2013; Perrone et al., 2002; Scharf and Colevas, 2006; Seruga et al., 2011; Sivendran et al., 2014). These data are summarized in Table 2, and individual characteristics of literature reviews are detailed in Table 3.

3.4. Main results of literature reviews

Among the 20 literature reviews that analyzed the overall quality of reporting, 8/20 (40%) highlighted an insufficient reporting of some items needed to evaluate the methodology of RCTs (Toulmonde et al., 2011; Grellety et al., 2014; Hui et al., 2012; Kober et al., 2006; Lai et al., 2006; Péron et al., 2012; Ziogas and Zintzaras, 2009) and OBS in oncology (Papathanasiou and Zintzaras, 2010). Some items such as randomization description, blinding details,

Table 2
Scores, tools and primary endpoints used in literature review studies (N = 58).

Tools of assessment	Literature review, n (%)
Reporting recommendations used	
CONSORT recommendations	22 (38)
ISOQOL recommendations	2 (3.4)
STROBE recommendations	1 (1.7)
Not detailed	33 (56.9)
Specific reporting score	
Overall Quality Score	7 (12.1)
Jadad scale	3 (5.2)
HRQOL outcomes in cancer clinical trials	2 (3.4)
Other	9 (15.5)
No score	37 (63.8)
Outcomes	
Overall quality of reporting	20 (34.5)
Health-Related Quality Of Life in RCTs	7 (12.1)
Primary endpoints in RCTs	6 (10.3)
Adverse events in phase II and phase III studies	5 (8.6)
Other	20 (34.5)

ISOQOL: International Society for Quality of Life Research; HRQOL: Health Related Quality of Life; RCTs: Randomised Clinical Trials.

allocation concealment and analysis by Intention-To-Treat (ITT) were noted as under-reported (Toulmonde et al., 2011; Ghimire et al., 2014; Lai et al., 2006; Lu et al., 2011; Péron et al., 2012). Three reviews studied the quality of reporting of abstracts describing RCTs in oncology (Ghimire et al., 2014; Hopewell and Clarke, 2005; Krzyzanowska et al., 2004), and concluded that the overall quality of reporting was still insufficient. Regarding the literature reviews assessing the quality of reporting for the primary endpoints (PEPs) in the RCTs, half of them lacked the precise definition of these PEPs (Le Tourneau et al., 2009; Mathoulin-Pelissier et al., 2008; Riechelmann et al., 2008). For the reporting of adverse events (AEs) in RCTs, most reviews of the literature concluded with a suboptimal reporting of AEs, with substantial selectivity and an heterogeneity of toxicity assessment between studies (Péron et al., 2013; Perrone et al., 2002; Scharf and Colevas, 2006; Sivendran et al., 2014). Concerning reporting on Health-Related Quality Of Life (HRQOL) in RCTs, the first two reviews showed insufficient reporting for the methodological description and the rationale for using specific measures of HRQOL (Efficace et al., 2004; Lee and Chi, 2000). However, three recent reviews of the literature have shown a significant improvement in the reporting of HRQOL studies in the last few years (Claassens et al., 2011; Efficace et al., 2007; Zikos et al., 2014). The main results are listed in Table 3.

3.5. Results of OQSR and missing items

The overall quality of reporting was performed for all literature reviews included in the study by using the Overall Quality Score for the literature Reviews (OQSR). The median OQSR was equal to 10 (interquartile range: 8–11). Five of the 58 literature reviews had a maximum OQSR (OQSR = 12/12, 9% of publications) (Dirven et al., 2014; Efficace et al., 2014a; Potter et al., 2011; Tanaka et al., 2012; Zikos et al., 2014). The minimum OQSR was equal to 4. Item 20 was the least frequently reported (results of each individual study: missing item for 42/58 reviews or 72% of the publications), followed by item 18 (individual study characteristics: missing item for 35/58 reviews or 60% of publications) and item 17 (study selection: missing item for 28/58 reviews or 48% of publications). (Table 3).

For the group of studies published between 1996 and 2009 (n = 29), mean OQSR was equal to 8 (SD: 2) and median OQSR was equal to 9 (interquartile range: 6–10). Literature reviews published between 2010 and 2014 have a significantly higher OQSR score: mean OQSR = 10 (SD: 2); median OQSR = 10 (interquartile range: 10–11), p = 0.0053 (Wilcoxon test). Results are detailed in Table 4.

Table 3
Literature review studies: Detailed characteristics, criteria analyzed for reporting evaluation, OQR results, Table 3 (continued).

Review literature	Number of analyzed studies	Type of analyzed studies (RCT/OBS)	Name of journal	Cancer type studied	Scores and tools used	Main criterion for judgment of reporting (MC)	Results on MC	OQR	PRISMA 2006–Missing items
Atkinson et al. (2011)	144	RCTs	Cancer	colon; rectum		Frequency of reporting of time-to-event outcomes and tumor response outcomes	OS and response rates: 99%, TTP: 44%, duration of response: 43%, PRS: 22%, and TTI failure: 12%	10	12; 17; 18; 19; 20
Bekeinan and Yehalom (2000)	61	RCTs	Int. J. Radiation Oncology Biol. Phys	lymphoma	RTDG procedure manual Philadelphia, 2005.	Reporting of six quality measures of Radiotherapy data	Target volume: 38%, radiation dose: 89%, fractionation: 64%, RT: 60%, RT: 60%, RT: 60%	10	12; 18; 19; 20
Blewone et al. (2012)	76	RCTs and OBS	Ann. Surg. Oncol.	esophagus		Reporting of rates of progression to surgical resection, inoperability at planned surgery following	Rate of inoperability at planned surgery: 31.6%, numbers of patients not progressing to surgery: 26.3%	9	10; 12; 18; 19; 20
Blewone et al. (2014)	62	RCTs	J. Natl. Cancer Inst.	esophagus; rectum; stomach; colon; liver metastases		Reporting of failure to complete neoadjuvant treatment, nonprogression after neoadjuvant treatment, and nonresection at planned surgery	Noncompletion of neoadjuvant treatment: 59.6%, nonprogression to surgery: 50%, nonresection at planned surgery: 37.1%	10	12; 18; 19; 20
Bottomley et al. (2003)	29	RCTs	J. Clin. Oncol.	lung		Reporting quality of the Health-Related Quality of Life (HRQL) trials	Trials exhibited good-quality research, weaknesses were: no clear hypothesis, no clear approach to presentation of results	10	13; 17
Burton and Altman (2004)	100	RCTs and OBS	Br. J. Cancer	all		Reporting of missing covariate data for prognostic models	Missing covariate data: 81%	4	8; 9; 11; 12; 13; 15; 16; 17; 18; 19; 20; 25
Byfeldt et al. (2015)	124	RCTs	Ann. Oncol.	all	CONSORT checklist, PROQOS	Reporting of Patient-Reported Outcomes (PROs)	PROs data: 38%, mean PROQOS score: 5.0 (range 0–11, 95% CI of the mean 4.4–5.5)	10	12; 18; 20
Cherry et al. (2009)	32	RCTs	J. Clin. Oncol.	all	CONSORT checklist, seven requirements for clinical clinical research, Helsinki Requirements	Reporting of ethical and methodological validity criteria	Helsinki requirements: poor compliance with requirements: poor compliance with requirements	10	15; 17; 22
Chassens et al. (2011)	53	RCTs	J. Clin. Oncol.	lung		Reporting quality of the Health-Related Quality of Life (HRQL) trials	The quality of HRQL reporting has improved: reporting of clinically significant differences and statistical testing of HRQL have improved.	10	13; 17
Droven et al. (2014)	14	RCTs	Eur. J. Cancer	brain	ISQOL checklist	Reporting of Patient-Reported Outcomes (PROs)	Complete data reporting (10 elements) (10): 11%, Premedication, growth factors support, dose adjustments for toxicities: <50%	12	9; 10; 12; 15; 18; 19; 20; 22; 25
Durl et al. (2010)	262	RCTs	J. Natl. Cancer Inst.	all	10 essential elements of RCT reporting (range: 0–10)	Reporting of important therapeutic details	Adjustments for toxicities: <50% of patients were treated. A lack of reporting on missing data, use of poorly validated tools, presentation and interpretation of the results.	5	9; 10; 12; 15; 18; 19; 20; 22; 25
Efficacy and Biomarker (2002)	5	RCTs	Eur. J. Cancer	brain	used practices in reporting Quality of Life (QOL)	Reporting of QOL assessment in RCTs	Level of reporting of Health-Related Quality of Life (HRQL) in randomized controlled trials (RCTs)	9	9; 11; 17
Efficacy et al. (2004)	31	RCTs	Eur. J. Cancer	colon; rectum		Level of reporting of Health-Related Quality of Life (HRQL) in randomized controlled trials (RCTs)	Level of reporting of Health-Related Quality of Life (HRQL) in randomized controlled trials (RCTs)	11	17
Efficacy et al. (2006)	44	RCTs	J. Clin. Epidemiol.	all	Minimum standard checklist for evaluating HRQL outcomes in cancer clinical trials (range 0–11)	Level of Patient-Reported Outcome (PRO) reporting in randomized controlled trials and alternative medicine trials	Level of Patient-Reported Outcome (PRO) reporting in randomized controlled trials and alternative medicine trials	11	17
Efficacy et al. (2007)	159	RCTs	Ann. Oncol.	prostate; breast; lung; colon; rectum	Minimum standard checklist for evaluating HRQL outcomes in cancer clinical trials (range 0–11)	Level of reporting of Health-Related Quality of Life (HRQL) in randomized controlled trials (RCTs)	Studies identified as being probably biased: 11% (range 0–11) between 1990 and 2000; 30.3% (range 0–11) between 2000 and 2007; 64.3% published after 2000; 64.3%	10	12; 17; 18; 19; 20

Table 3 (Continued)

Review literature	Number of analyzed studies (RCTs/OBS)	Type of studies	Name of journal	Cancer type studied	Scores and tools used	Main criterion for judgment of reporting (MC)	Results on MC	OQR	PRISMA 2009-Missing items
Efficace et al. (2014a)	50	RCTs	Eur. J. Cancer	cervix; ovary; endometrium; vagina; vulva	ISQOL consensus-based recommendations for reporting of PRIs in RCT publications	Level of Patient-Reported Outcome (PRO) reporting in RCTs in the gynecological cancer sites	16 RCTs (32%) were considered to be of high-quality and thus able to inform clinical decision making	12	18; 20
Efficace et al. (2014b)	65	RCTs	Eur. J. Oncol.	prostate	ISQOL consensus-based recommendations for reporting of PRIs in RCT publications	PRO) reporting in RCTs of prostate cancer	Mean OQS: 8.2 (range: 5–13; 95%CI: 8.0–8.3) for the overall period and 9.0 (range: 5–13; 95%CI: 8.7–9.2) in the post-CONSORT period.	11	15; 20; 22
Ghimire et al. (2014)	956	RCTs	J. Clin. Epidemiol.	all	Overall Quality Score (OQS; range 0–18) for abstracts, CONSORT checklist	Overall quality of reporting of oncology RCT abstracts	Median OQS: 28 (range: 9–35); Primary end point definition: 68.6%; Secondary end point definition: 68.6%; Jadad score (Mean = 1.945; SD = 1.01); KMI score (Mean = 0.843; SD = 0.07); adherence to CONSORT (Mean = 10.9/137).	10	8; 15; 17; 18; 20; 22
Griffithy et al. (2014)	156	RCTs	Ann. Oncol.	all	Overall Quality Score (OQS; range 0–44), CONSORT checklist, GROSS	Overall quality of reporting of phase II trials	Median OQS: 28 (range: 9–35); Primary end point definition: 68.6%; Secondary end point definition: 68.6%; Jadad score (Mean = 1.945; SD = 1.01); KMI score (Mean = 0.843; SD = 0.07); adherence to CONSORT (Mean = 10.9/137).	9	8; 15; 17; 18; 20; 22
Guo et al. (2014)	227	RCTs	Nurs. Res.	all	CONSORT checklist (range 0–37), Jadad score (range 0–5), key methodologic index (KMI; range 0–3)	Overall quality of reporting in published cancer nursing RCTs	Method of blinding: 12% of abstracts, a reported ITT analysis (10%), and 13% in 2002. Lack of allocation concealment reporting.	11	18; 20; 22
Hopewell and Clarke (2005)	484	RCTs	Clin. Trials	all	CONSORT checklist and guidance on the structured reporting of journal abstracts and short reports	Overall quality of reporting in abstracts describing RCTs presented at ASCO meetings (1992 and 2002)	Method of blinding: 12% of abstracts, a reported ITT analysis (10%), and 13% in 2002. Lack of allocation concealment reporting.	6	9; 10; 15; 17; 18; 20; 22; 25
Hui et al. (2012)	44	RCTs	Support. Care Cancer	all	Overall Quality Score (OQS; range 0–15), key methodologic index (KMI; range 0–3), Jadad score, CONSORT	Overall quality of reporting in palliative and supportive cancer RCTs	Median CONSORT OQS was 9 (interquartile range 7–11). KMI was 0.7 (range 0–3) and Jadad score was 2 (interquartile range 1–3).	10	8; 10; 17; 18; 20
Kober et al. (2006)	242	RCTs	J. Natl. Cancer Inst.	lymphoma	CONSORT checklist (14 items)	Overall quality of reporting in published RCTs involving lymphoma	At least 6/14 items addressed in the studies. 75% necessary to the study; 20% of the studies.	11	18; 20
Krzyzanowska et al. (2004)	510	RCTs	J. Clin. Oncol.	all	Quality Score (range 0–10)	Overall quality of reporting in abstracts describing RCTs included in the primary treatment of brain tumors	Explicit identification of the primary end point: 22%. Median quality score: 5.5.	11	17; 18; 20
Lai et al. (2006)	74	RCTs	J. Clin. Oncol.	brain	CONSORT checklist (15 items), Key Methodologic Factors (3 items)	Overall reporting quality of RCTs in the primary treatment of brain tumors	At least 8/15 methodologic items reported: <50% of trials. All the key methodologic factors reported.	9	11; 15; 18; 20; 22
Le Faour et al. (2009)	40	RCTs	J. Clin. Oncol.	OBL	CONSORT checklist (76 items)	Reporting of time-to-event end points and the tracking of failures	17 of 46 end points reported. Locoregional control and overall survival: 70% of primary end points.	6	12; 13; 15; 17; 19; 22; 25
Lee and Chi (2000)	20	RCTs	J. Clin. Epidemiol.	all	CONSORT checklist (71 items) and Quality of Life Reporting (QLR) checklist (76 items)	Reporting of Health-Related Quality of Life (HRQL) in clinical cancer trials	Number of items in QLR checklist: 42% (range 15–85%). median number of items in QLR checklist: 20.	11	17; 20
Lu et al. (2011)	46	RCTs	Expert Rev. Anticancer Ther.	all	CONSORT checklist	Overall quality of reporting in RCTs on the treatment of cancer pain	CONSORT checklist: 70% (range 62–96%).	7	14; 15; 17; 18; 20; 22
Lu et al. (2013)	32	RCTs	Expert Rev. Anticancer Ther.	all	Overall Quality Score (OQS), Score (MS), CONSORT checklist and STRCTA	Overall quality of reporting in RCTs on acupuncture for cancer pain.	21.7% approval by an ethics committee: 32.8% Statement: 4.0 median OQS based on STRCTA: 0	11	17; 18; 20; 22

Author (Year)	Study ID	Journal	Population	Intervention	Outcome	Quality Score	Comments
Murphy et al. (2005)	208	Cancer	colon, rectum	RCTs and OBS	10 Prognostic factors	4	Reporting of prognostic factors in the Meta-analysis of Cochrane Cancer trials from chemotherapy and surgery trials
Mohrlein-Pelissier et al. (2008)	125	J. Clin. Oncol.	all	RCTs	CONSORT checklist	8	The median score: 7 and the mean score: 6. Accurate information about analysis: 32%.
Mhaskar et al. (2012)	429	J. Clin. Epidemiol.	all	RCTs		10	Overall reporting of methodological quality was poor and did not reflect the methodological quality of RCTs.
Mitchell et al. (2009)	197	Breast Cancer Res. Treat. Epidemiol.	breast	RCTs	Reporting of analysis of outcomes by race/ethnicity in breast cancer RCTs	9	Accrual reported by race: 17%.
Paik, Wenzel and Zujewski (2010)	244	OBS	all	STROBE checklist (25 items/sub-items)	Overall reporting quality of observational studies concerning cancer	10	Overall reporting quality of observational studies concerning cancer reported: 7/9%.
Pat et al. (2008)	45	Lung Cancer	lung	RCTs	Weighted CONSORT score (range 0–31), CONSORT checklist	9	Median weighted CONSORT score: 25/31 (range: 12–30)
Petron et al. (2012)	357	J. Natl. Cancer Inst.	all	RCTs	2001 Overall Quality Score (OQS: range 0–18 and 2010 OQS (range 0–27), CONSORT checklist	11	Mean 2001 OQS: 13.4, mean 2010 OQS: 19.3, method used for the random allocation: 29%, whether stratified: 41%
Petron et al. (2013)	325	J. Clin. Oncol.	all	RCTs	AERQS (range: 0–16), CONSORT checklist	9	Mean AERQS: 10.1/16 points. Methodology of AE collection: 10%.
Perzoug et al. (2002)	122	J. Clin. Oncol.	breast	RCTs	Reporting of toxicity in phase II chemotherapy trials in breast cancer	10	Scoring to weight overall: 15%. Stratification of patients: 100%. 34.4% Planned modalities for assessment of toxicity: 79.5%.
Potter et al. (2011)	134	J. Natl. Cancer Inst.	breast	RCTs and OBS	Cochrane Risk of Bias tool	12	Reporting of surgical outcomes in breast reconstruction
Rieschmann et al. (2008)	125	J. Clin. Oncol.	all	RCTs	Reporting of Primary End Points (PEP) in phase II cancer trials	7	Adequacy between methods and results in the numbers of complications: 47%. Complications defined: 84.9% and graded: 38.2%. Rate used as PEP: 58%.
Scharf and Colevas (2006)	22	J. Clin. Oncol.	all	RCTs	NCI Clinical Data Update System (CDUS)	5	High-grade AEs matched to agent-attributable AEs in CDUS: 17: 18; 19: 20; 22: 25.
Seruga et al. (2011)	36	J. Clin. Oncol.	all	RCTs	Comparison of Adverse Events (AE) of treatments identified from updated labels, with those reported in the original product literature: 72%	7	30/16 (39%) of all serious AEs and 15/38 (39%) of potentially fatal AEs not described in any published literature.
Shirazi et al. (2012)	67	Gastric Cancer	stomach	RCTs	Reporting of patients characteristics and stratification factors that have been used	7	Histology, number of disease sites, and adjuvant treatment were reported in less than 50% of trials. Stratification was conducted in 60%.
Svendran et al. (2014)	175	J. Clin. Oncol.	all	RCTs	14 adverse event-reporting elements (range: 0–14), CONSORT checklist	10	Median completeness score: 8/14 (range: 3–12). Only AEs occurring above a threshold rate or severity: 96%.

Table 3 (Continued)

Review literature	Number of analyzed studies (RCTs/OBS)	Type of analyzed studies	Name of journal	Cancer type studied	Scores and tools used	Main criterion for judgment of reporting (MC)	Results on MC	OQSR	PRISMA 2009-Missing items
Sources et al. (2004)	56	RCTs	BMJ	all		Correlation between poor reporting and poor methods	Adequate allocation concealment: 100% (n=56); intention to treat analysis: 83%	6	10; 13; 15; 18; 20; 22; 25
Sorbye et al. (2007)	143	RCTs	Ann. Oncol.	colon; rectum		Reporting of patient characteristics and identification of stratification factors in phase I and phase III studies	Laboratory values (PAC, LDH and WB cell count) reported: 5–9%. Study factors with any consistency	5	8; 9; 10; 11; 12; 15; 17; 18; 19; 20; 22; 25
Sitt et al. (2008)	33	RCTs	Hematol Oncol Stem Cell Transf. Cancer	all	CONSORT checklist (19 items)	Overall quality of reporting of RCTs	Average adherence to 19 items of the CONSORT checklist: 79.3% (95% CI)	9	8; 9; 10; 18; 20; 22
Swanson and Bailar (2002)	261	RCTs and OBS		all		Reporting of women and minorities in cancer treatment and prevention trials	Age and gender of study participants: more than 90% race or ethnicity of study participants: more than 30%	5	9; 12; 13; 15; 17; 18; 19; 20; 22; 25
Tanaka et al. (2012)	72	RCTs	Chn. Cancer Res.	all	CONSORT checklist (11 items)	Overall quality of reporting of noninferiority trials in cancer	Consistency when results are not statistically significant: 9.7% or when noninferiority margin was not defined: 6.9%	12	
Thomas et al. (2004)	393	RCTs	Eur. J. Cancer	all		Evolution in statistical design reporting for phase II cancer clinical trials	2000. Information on statistical designs used: 15% in 1995 and 46% in 2000	10	15; 17; 18; 20; 22
Thomas et al. (2009)	137	RCTs	Med. Law	breast	Jadad scale (range 0–13); Bardenheier scale (range 0–1)	Overall quality of reporting and ethical quality of phase III breast cancer RCTs	Mean quality of reporting was 9.88 (s.d. 1.43, Jadad score <9 for 49 RCT (35.8%). Mean ethical quality was 0.457 (s.d. 0.17, range 0–1), median BMS: 1.0 (range: 0–2)	6	7; 8; 9; 15; 17; 18; 19; 20; 22; 25
Toulmeade et al. (2011)	72	RCTs	J. Clin. Oncol.	sarcoma	Overall Quality Score (OQS; range 0–15), Methodologic Index Score (MIS; range 0–3), CONSORT checklist	Overall quality of reporting in sarcoma RCTs	Sample size calculations: 68%, who were reported: 88%. Most (80%) many individuals were screened: 36%	10	15; 18; 20; 22
Treweek (2011)	207	RCTs	J. Clin. Epidemiol.	breast	Jadad scale (range 0–13); Bardenheier scale (range 0–1)	Reporting of key items of patient information in trial reports	Mean Jadad score: 9.86 (s.d. 1.117, range 0–13), mean BMS score: 1.0 (range 0–2), mean BMS: 1.0 (range: 0–2)	10	15; 22
Tuohi et al. (2005)	231	RCTs	J. Med. Ethics	all	Jadad scale (range 0–13); Bardenheier scale (range 0–1)	Overall quality of reporting and ethical quality of phase III cancer trials	The mean Bardenheier score was 0.42 (s.d. 0.133)	8	7; 8; 9; 15; 18; 19; 20; 22
Vera-Radillo et al. (2013)	164	RCTs	Ann. Oncol.	breast	Toxicity hierarchy scale (range: 0–7)	Reporting of the primary end point and safety in RCTs for breast cancer	Reporting of the PE not biased: 67%. Reporting of toxicity not biased: 33%	11	18; 20
You et al. (2012)	366	RCTs	J. Clin. Oncol.	all	CONSORT checklist	Concordance of Primary End Points between publication reports and the corresponding registry	PEP differed between registration and publication: 14% trials with concordance; 86% trials with discordance in RCTs	11	18; 20
Zhao et al. (2014)	30	RCTs	Lancet Oncol.	lung		Adequacy of reporting of Health-Related Quality-Of-Life (HRQL) methods in RCTs of small cell lung cancer	HRQL domains covered by the questionnaires considered as adequate: 80%. Most studies did not use HRQL research hypothesis.	12	
Zingales and Zinzaras (2009)	261	RCTs	Ann. Epidemiol.	leukemia; myelodysplastic syndromes	CONSORT checklist (24 items)	Overall reporting quality of published RCTs concerning myeloid hematologic malignancies	13 of the 24 items were addressed in >75% of the studies. Most (80%) were reported in >50% of the studies	11	20

RCT: Randomized Controlled Trial; OS: Overall Survival; TTP: Time-To-Progression; PFS: Progression-Free Survival; ITT: Intention To Treat; STRUTIA: Standards for Reporting Interventions in Controlled Trials of Acupuncture; PAC: Alkaline phosphatase; LDH: Lactate dehydrogenase; WB: white blood

Table 4
Mean and median OQSR for literature reviews published before (<2010) or after publication of PRISMA statement.

	Number of studies	Mean OQSR (SD)	Median OQSR [interquartile] (min-max)	P-value
1996–2009	29	8 (2)	9 [6–10] (4–11)	0,0053 ^a
2010–2014	29	10 (2)	10 [10,11] (5–12)	

^a Wilcoxon test; OQSR: Overall Quality Score for literature Reviews; SD: Standard Deviation; min: minimum OQSR value; max: maximum OQSR value.

4. Discussion

To our knowledge, this literature review is the first that summarizes the current state of research on the quality of reporting of randomized controlled trials (RCTs) and observational studies (OBS) in oncology. A great part of reviews in our study rated the overall quality of reporting of RCTs in oncology (Toulmonde et al., 2011; Ghimire et al., 2014; Grellety et al., 2014; Guo et al., 2014; Hopewell and Clarke, 2005; Hui et al., 2012; Kober et al., 2006; Krzyzanowska et al., 2004; Lai et al., 2006; Lu et al., 2011; Lu et al., 2013; Pat et al., 2008; Péron et al., 2012; Soares et al., 2004; Süt et al., 2008; Tanaka et al., 2012; Thoma et al., 2009; Tuech et al., 2005; Ziogas and Zintzaras, 2009), highlighting an inconsistent quality of reporting for methodological items which are however essential for an informed interpretation of the results: The omission in the manuscripts of items as important as the description of randomization, blinding details, allocation concealment and analysis by intention to treat should be criteria for publication refusal for medical journals editors. One can also notice that no reviews specifically analyzed the quality of reporting of statistical analysis, especially on the tools used and the rationale for using these tools. There is an item called “statistical methods” in CONSORT (Schulz et al., 2010) and STROBE (Vandenbroucke et al., 2007) statements, the accurate analysis of this aspect of reporting seems important because the validity and robustness of the results obtained in RCTs and OBS depend in part on the type of statistical analysis performed.

The second point to underline is that the quality of reporting of OBS in oncology has not been widely assessed, with only one study (2% of publications) that addressed this issue (Papathanasiou and Zintzaras, 2010) and with 5 studies that examined both the reporting of RCTs and OS (9% of publications) (Blencowe et al., 2012; Burton and Altman, 2004; Martin et al., 2006; Potter et al., 2011; Swanson and Bailar, 2002). In the review of Papathanasiou A et al. (Papathanasiou and Zintzaras, 2010), the authors remind us that a substantial amount of clinical and public health knowledge originates from OBS (Glasziou et al., 2004), so it seems essential to widely assess the quality of reporting of these studies in the oncology setting.

Although 50% of reviews were published before or during the year of publication of the PRISMA statement, the intrinsic quality of reporting of the included reviews is quite good: The median OQSR is equal to 10 and 5 reviews have a maximum OQSR (12/12) (Dirven et al., 2014; Efficace et al., 2014a; Potter et al., 2011; Tanaka et al., 2012; Zikos et al., 2014). There has also been a trend towards the improvement in the overall quality of the review reporting since the publication of PRISMA statement in 2009: median OQSR_{1996–2009} = 9 (interquartile range: 6–10) vs median OQSR_{2010–2014} = 10 (interquartile range: 10–11), $p=0.0053$. The growing adherence of medical journal editors to the PRISMA statement is an essential prerequisite for improving the overall quality of reporting of reviews and meta analyzes, as well as adherence to the CONSORT statement for RCTs and the STROBE statement for OBS. However, it should be noted that the reporting of individual study results and individual study characteristics in the reviews is low (item 20, reported in 28% of the publications and item 18, reported in 40% of the publications). It would be interesting to have a sys-

tematic access (via an Appendix or a web link) for characteristics and individual results of RCTs and OBS analyzed in the reviews, and especially in the reporting field: This could allow readers to make their own judgments about the relevance of included studies. Reporting of study selection can also be considered insufficient in this study (item 17, reported in 52% of publications). As explained in the PRISMA statement (Moher et al., 2009), adequate reporting of this item is particularly important because literature which was identified primarily from references or experts may be prone to citation or publication bias.

More than half of the literature reviews have examined the quality of reporting for trials, whatever the primary cancer studied (31/58 reviews, 53%). This implies that, for the majority of the authors of literature reviews, the primary site of the cancer does not appear to be a factor that can influence the quality of reporting. But to date, no scientific evidence can confirm this argument.

The majority of those literature reviews have been published in medical journals in Oncology and with a high to very high impact factor (*Journal of Clinical Oncology*, *Annals of Oncology*, *Journal of the National Cancer Institute* or *European Journal of Cancer*, 53% of cases). The publication of these reviews into highly reed journals, could be a proxy to assess the peer review system quality: Even if published into those Journals, where peer review is supposed to be more thorough than elsewhere, some mandatory information is lacking, even if the overall quality level remains high.

We should also remember that the overall quality of reporting for RCTs does not always reflect the intrinsic quality of the corresponding research protocols. In the study of Soares et al. (2004), the authors demonstrate that bad reporting does not mean bad methods for randomized trials. These conclusions are confirmed by the study of Mhaskar et al. (2012), which also underlines the importance of publication of RCT protocols in the public domain.

The current review has a number of limitations. First, our literature review was limited to published studies, and therefore it is potentially subject to publication bias. However, we can assume that literature reviews that had bad quality reporting were rejected for publication. The inclusion of such reviews in our study would have decreased the median OQSR and this could have led to less reliable results. Overall Quality of reporting for literature reviews was quantified according to a score (OQSR) that has not been used so far, but there is currently no validated score based on the PRISMA statement.

In conclusion, this review of systematic reviews has confirmed the existence of insufficient reporting of methodological items in RCTs. Furthermore, the OBS reporting quality has been little evaluated till now. The specific evaluation of the quality of reporting for statistical tests has also been little analyzed and a literature review is necessary to specifically assess this reporting element. The intrinsic quality of reporting of included reviews was considered satisfactory in our study and with a significant improvement since the publication of the PRISMA statement.

Contributors

Study concept: Franck Chauvin, Nicolas Magné.
Study Design: Romain Rivoirard, Aurélie Bourmaud.

Data acquisition: Romain Rivoirard.
 Data analysis and interpretation: Romain Rivoirard, Aurélie Bourmaud, Alexis Vallard.
 Statistical analysis: Mathieu Oriol, Fabien Tinquaut.
 Quality control of data: Benoîte Mery, Julien Langrand-Escure.
 Manuscript preparation: Romain Rivoirard, Aurélie Bourmaud, Alexis Vallard, Nicolas Magné.
 Manuscript review: Pierre Fournel, Alexis Vallard, Nicolas Magné, Aurélie Bourmaud.

Funding/Support

No funding.

Conflict of interest statement

None.

Acknowledgment

The authors would like to thank Susan Guillaumond (Activenglish) for translation and editing services.

References

- A proposal for structured reporting of randomized controlled trials, 1994. The standards of reporting trials group. *JAMA* 272, 1926–1931.
- Arkenau, H.-T., Nordman, I., Dobbins, T., Ward, R., 2011. Reporting time-to-event endpoints and response rates in 4 decades of randomized controlled trials in advanced colorectal cancer. *Cancer* 117, 832–840.
- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al., 1996. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 276, 637–639.
- Bekelman, J.E., Yahalom, J., 2009. Quality of radiotherapy reporting in randomized controlled trials of Hodgkin's lymphoma and non-Hodgkin's lymphoma: a systematic review. *Int. J. Radiat. Oncol. Biol. Phys.* 73, 492–498.
- Black, N., 1996. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 312, 1215–1218.
- Blencowe, N.S., McNair, A.G.K., Davis, C.R., Brookes, S.T., Blazeby, J.M., 2012. Standards of outcome reporting in surgical oncology: a case study in esophageal cancer. *Ann. Surg. Oncol.* 19, 4012–4018.
- Blencowe, N.S., Chana, P., Whistance, R.N., Stevens, D., Wong N.A., C.S., Falk, S.J., et al., 2014. Outcome reporting in neoadjuvant surgical trials: a systematic review of the literature and proposals for new standards. *J. Natl. Cancer Inst.* 106.
- Bottomley, A., Efficace, F., Thomas, R., Vanvoorden, V., Ahmedzai, S.H., 2003. Health-related quality of life in non-small-cell lung cancer: methodologic issues in randomized controlled trials. *J. Clin. Oncol.* 21, 2982–2992.
- Burton, A., Altman, D.G., 2004. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. *Br. J. Cancer* 91, 4–8.
- Bylicki, O., Gan, H.K., Joly, F., Mailet, D., You, B., Péron, J., 2015. Poor patient-reported outcomes reporting according to CONSORT guidelines in randomized clinical trials evaluating systemic cancer therapy. *Ann. Oncol.* 26, 231–237.
- Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature, 1994. Working group on recommendations for reporting of clinical trials in the biomedical literature. *Ann. Intern. Med.* 121, 894–895.
- Les cancers en France en 2013. Collection état des lieux et des connaissances, ouvrage collectif édité par I (Les cancers en France en 2013. Collection état des lieux et des connaissances, ouvrage collectif édité par l'INCa, Boulogne-Billancourt janvier 2014 [in French]).
- Cherny, N.I., Abernethy, A.P., Strasser, F., Sapir, R., Currow, D., Zafar, S.Y., 2009. Improving the methodologic and ethical validity of best supportive care studies in oncology: lessons from a systematic review. *J. Clin. Oncol.* 27, 5476–5486.
- Claessens, L., van Meerbeeck, J., Coens, C., Quinten, C., Ghislain, I., Sloan, E.K., et al., 2011. Health-related quality of life in non-small-cell lung cancer: an update of a systematic review on methodologic issues in randomized controlled trials. *J. Clin. Oncol.* 29, 2104–2120.
- Dirven, L., Taphoorn, M.J.B., Reijneveld, J.C., Blazeby, J., Jacobs, M., Pusic, A., et al., 2014. The level of patient-reported outcome reporting in randomised controlled trials of brain tumour patients: a systematic review. *Eur. J. Cancer* 50, 2432–2448.
- Duff, J.M., Leather, H., Walden, E.O., LaPlant, K.D., George, T.J., 2010. Adequacy of published oncology randomized controlled trials to provide therapeutic details needed for clinical application. *J. Natl. Cancer Inst.* 102, 702–705.
- Efficace, F., Bottomley, A., 2002. Health related quality of life assessment methodology and reported outcomes in randomised controlled trials of primary brain cancer patients. *Eur. J. Cancer* 38, 1824–1831.
- Efficace, F., Bottomley, A., Vanvoorden, V., Blazeby, J.M., 2004. Methodological issues in assessing health-related quality of life of colorectal cancer patients in randomised controlled trials. *Eur. J. Cancer* 40, 187–197.
- Efficace, F., Horneber, M., Lejeune, S., Van Dam, F., Leering, S., Rottmann, M., et al., 2006. Methodological quality of patient-reported outcome research was low in complementary and alternative medicine in oncology. *J. Clin. Epidemiol.* 59, 1257–1265.
- Efficace, F., Osoba, D., Gotay, C., Sprangers, M., Coens, C., Bottomley, A., 2007. Has the quality of health related quality of life reporting in cancer clinical trials improved over time?: Towards bridging the gap with clinical decision making. *Ann. Oncol.* 18, 775–781.
- Efficace, F., Jacobs, M., Pusic, A., Greimel, E., Picciocchi, A., Kieffer, J.M., et al., 2014a. Patient-reported outcomes in randomised controlled trials of gynaecological cancers: investigating methodological quality and impact on clinical decision-making. *Eur. J. Cancer* 50, 1925–1941.
- Efficace, F., Feuerstein, M., Fayers, P., Cafaro, V., Eastham, J., Pusic, A., et al., 2014b. Patient-reported outcomes in randomised controlled trials of prostate cancer: methodological quality and impact on clinical decision making. *Eur. Urol.* 66, 416–427.
- Ghimire, S., Kyung, E., Lee, H., Kim, E., 2014. Oncology trial abstracts showed suboptimal improvement in reporting: a comparative before-and-after evaluation using CONSORT for Abstract guidelines. *J. Clin. Epidemiol.* 67, 658–666.
- Glasziou, P., Vandenbroucke, J.P., Chalmers, I., 2004. Assessing the quality of research. *BMJ* 328, 39–41.
- Grellety, T., Petit-Monéger, A., Diallo, A., Mathoulin-Pelissier, S., Italiano, A., 2014. Quality of reporting of phase II trials: a focus on highly ranked oncology journals. *Ann. Oncol.* 25, 536–541.
- Guo, J.-W., Sward, K.A., Beck, S.L., Staggers, N., 2014. Quality of reporting randomized controlled trials in cancer nursing research. *Nurs. Res.* 63, 26–35.
- Guyatt, G.H., Oxman, A.D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., et al., 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336, 924–926.
- Hopewell, S., Clarke, M., 2005. Abstracts presented at the American Society of Clinical Oncology conference: how completely are trials reported? *Clin. Trials* 2, 265–268.
- Hopewell, S., Collins, G.S., Boutron, I., Yu, L.-M., Cook, J., Shanyinde, M., et al., 2014. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. *BMJ* 349, g4145.
- Hui, D., Arthur, J., Dalal, S., Bruera, E., 2012. Quality of the supportive and palliative oncology literature: a focused analysis on randomized controlled trials. *Support. Care Cancer* 20, 1779–1785.
- International Committee of Medical Journal Editors, 2013. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals [cited 13/12/2014] Available from URL: www.icmje.org.
- Ivers, N.M., Taljaard, M., Dixon, S., Bennett, C., McRae, A., Taleban, J., et al., 2011. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000–8. *BMJ* 343, d5886.
- Kober, T., Trelle, S., Engert, A., 2006. Reporting of randomized controlled trials in Hodgkin lymphoma in biomedical journals. *J. Natl. Cancer Inst.* 98, 620–625.
- Krzyzanowska, M.K., Pintilie, M., Brezden-Masley, C., Dent, R., Tannock, I.F., 2004. Quality of abstracts describing randomized trials in the proceedings of American Society of Clinical Oncology meetings: guidelines for improved reporting. *J. Clin. Oncol.* 22, 1993–1999.
- Lai, R., Chu, R., Fraumeni, M., Thabane, L., 2006. Quality of randomized controlled trials reporting in the primary treatment of brain tumors. *J. Clin. Oncol.* 24, 1136–1144.
- Le Tourneau, C., Michiels, S., Gan, H.K., Siu, L.L., 2009. Reporting of time-to-event end points and tracking of failures in randomized trials of radiotherapy with or without any concomitant anticancer agent for locally advanced head and neck cancer. *J. Clin. Oncol.* 27, 5965–5971.
- Lee, C.W., Chi, K.N., 2000. The standard of reporting of health-related quality of life in clinical cancer trials. *J. Clin. Epidemiol.* 53, 451–458.
- Lu, L., Zeng, J., Chen, Y., 2011. Quality of reporting in randomized controlled trials conducted in China on the treatment of cancer pain. *Expert Rev. Anticancer Ther.* 11, 871–877.
- Lu, L., Liao, M., Zeng, J., He, J., 2013. Quality of reporting and its correlates among randomized controlled trials on acupuncture for cancer pain: application of the CONSORT 2010 Statement and STRICTA. *Expert Rev. Anticancer Ther.* 13, 489–498.
- Martin, R.C.G., Augenstein, V.A., Scoggins, C.R., McMasters, K.M., 2006. Quality of survival reporting in chemotherapy and surgery trials in patients with metastatic colorectal carcinoma. *Cancer* 106, 1389–1394.
- Mathoulin-Pelissier, S., Gourgou-Bourgade, S., Bonnetain, F., Kramar, A., 2008. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J. Clin. Oncol.* 26, 3721–3726.
- Mhaskar, R., Djulbegovic, B., Magazín, A., Soares, H.P., Kumar, A., 2012. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *J. Clin. Epidemiol.* 65, 602–609.
- Mitchell, K.W., Carey, L.A., Peppercorn, J., 2009. Reporting of race and ethnicity in breast cancer research: room for improvement. *Breast Cancer Res. Treat.* 118, 511–517.

- Moher, D., Pham, B., Jones, A., Cook, D.J., Jadad, A.R., Moher, M., et al., 1998. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 352, 609–613.
- Moher, D., Schulz, K.F., Altman, D.G., 2001. CONSORT: The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Med. Res. Methodol.* 1, 2.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6, e1000097.
- Péron, J., Pond, G.R., Gan, H.K., Chen, E.X., Almufti, R., Mailet, D., et al., 2012. Quality of reporting of modern randomized controlled trials in medical oncology: a systematic review. *J. Natl. Cancer Inst.* 104, 982–989.
- Péron, J., Mailet, D., Gan, H.K., Chen, E.X., You, B., 2013. Adherence to CONSORT adverse event reporting guidelines in randomized clinical trials evaluating systemic cancer therapy: a systematic review. *J. Clin. Oncol.* 31, 3957–3963.
- Papathanasiou, A.A., Zintzaras, E., 2010. Assessing the quality of reporting of observational studies in cancer. *Ann. Epidemiol.* 20, 67–73.
- Pat, K., Dooms, C., Vansteenkiste, J., 2008. Systematic review of symptom control and quality of life in studies on chemotherapy for advanced non-small cell lung cancer: how CONSORTed are the data? *Lung Cancer* 62, 126–138.
- Perrone, F., De Maio, E., Maione, P., Di Maio, M., Ottaiano, A., Pensabene, M., et al., 2002. Survey of modalities of toxicity assessment and reporting in noncomparative prospective studies of chemotherapy in breast cancer. *J. Clin. Oncol.* 20, 52–57.
- Potter, S., Brigic, A., Whiting, P.F., Cawthorn, S.J., Avery, K.N.L., Donovan, J.L., et al., 2011. Reporting clinical outcomes of breast reconstruction: a systematic review. *J. Natl. Cancer Inst.* 103, 31–46.
- Rennie, D., 1992. Suspended judgment. Editorial peer review: let us put it on trial. *Control. Clin. Trials* 13, 443–445.
- Rennie, D., 2001. CONSORT revised—improving the reporting of randomized trials. *JAMA* 285, 2006–2007.
- Riechelmann, R.P., Dounaevskaia, V., Krzyzanowska, M.K., 2008. Quality of reporting primary outcomes in phase II cancer trials. *J. Clin. Oncol.* 26, 5486–5488.
- Süt, N., Senocak, M., Uysal, O., Köksalan, H., 2008. Assessing the quality of randomized controlled trials from two leading cancer journals using the CONSORT statement. *Hematol. Oncol. Stem Cell Ther.* 1, 38–43.
- Scharf, O., Colevas, A.D., 2006. Adverse event reporting in publications compared with sponsor database for cancer clinical trials. *J. Clin. Oncol.* 24, 3933–3938.
- Schulz, K.F., Altman, D.G., Moher, D., 2010. CONSORT Group: CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 340, c332.
- Seruga, B., Sterling, L., Wang, L., Tannock, I.F., 2011. Reporting of serious adverse drug reactions of targeted anticancer agents in pivotal phase III clinical trials. *J. Clin. Oncol.* 29, 174–185.
- Shitara, K., Ikeda, J., Kondo, C., Takahari, D., Ura, T., Muro, K., et al., 2012. Reporting patient characteristics and stratification factors in randomized trials of systemic chemotherapy for advanced gastric cancer. *Gastric Cancer* 15, 137–143.
- Sivendran, S., Latif, A., McBride, R.B., Stensland, K.D., Wisnivesky, J., Haines, L., et al., 2014. Adverse event reporting in cancer clinical trial publications. *J. Clin. Oncol.* 32, 83–89.
- Soares, H.P., Daniels, S., Kumar, A., Clarke, M., Scott, C., Swann, S., et al., 2004. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 328, 22–24.
- Sorbye, H., Köhne, C.-H., Sargent, D.J., Glimelius, B., 2007. Patient characteristics and stratification in medical treatment studies for metastatic colorectal cancer: a proposal for standardization of patient characteristic reporting and stratification. *Ann. Oncol.* 18, 1666–1672.
- Sorensen, A.A., Wojahn, R.D., Manske, M.C., Calfee, R.P., 2013. Using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement to assess reporting of observational trials in hand surgery. *J. Hand Surg. Am.* 38, 1584–1589, e2.
- Swanson, G.M., Bailar, J.C., 2002. Selection and description of cancer clinical trials participants?science or happenstance? *Cancer* 95, 950–959.
- Tanaka, S., Kinjo, Y., Kataoka, Y., Yoshimura, K., Teramukai, S., 2012. Statistical issues and recommendations for noninferiority trials in oncology: a systematic review. *Clin. Cancer Res.* 18, 1837–1847.
- Thezenas, S., Duffour, J., Culine, S., Kramar, A., 2004. Five-year change in statistical designs of phase II trials published in leading cancer journals. *Eur. J. Cancer* 40, 1244–1249.
- Thoma, V., Bridoux, V., Lefebvre, B., Wattiez, A., Nisand, I., Tuech, J.-J., 2009. Methodological and ethical quality in phase III breast cancer trials. *Med. Law* 28, 637–648.
- Toulmonde, M., Bellera, C., Mathoulin-Pelissier, S., Debled, M., Bui, B., Italiano, A., 2011. Quality of randomized controlled trials reporting in the treatment of sarcomas. *J. Clin. Oncol.* 29, 1204–1209.
- Treweek, S., 2011. Incomplete reporting of recruitment information in breast cancer trials published between 2003 and 2008. *J. Clin. Epidemiol.* 64, 1216–1222.
- Tuech, J.J., Pessaux, P., Moutel, G., Thoma, V., Schraub, S., Herve, C., 2005. Methodological quality and reporting of ethical requirements in phase III cancer trials. *J. Med. Ethics* 31, 251–255.
- Vandenbroucke, J.P., von Elm, E., Altman, D.G., Gøtzsche, P.C., Mulrow, C.D., Pocock, S.J., et al., 2007. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 4, e297.
- Vera-Badillo, F.E., Shapiro, R., Ocana, A., Amir, E., Tannock, I.F., 2013. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. *Ann. Oncol.* 24, 1238–1244.
- Von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P., et al., 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med.* 4, e296.
- World Health Organization, International Agency for Research on Cancer, 2017. GLOBOCAN2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. Cancer Fact Sheets [cited 12/12/2014]. Available from URL: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
- You, B., Gan, H.K., Pond, G., Chen, E.X., 2012. Consistency in the analysis and reporting of primary end points in oncology randomized controlled trials from registration to publication: a systematic review. *J. Clin. Oncol.* 30, 210–216.
- Zikos, E., Ghislain, I., Coens, C., Ediebah, D.E., Sloan, E., Quinten, C., et al., 2014. Health-related quality of life in small-cell lung cancer: a systematic review on reporting of methods and clinical issues in randomised controlled trials. *Lancet Oncol.* 15, e78–89.
- Ziogas, D.C., Zintzaras, E., 2009. Analysis of the quality of reporting of randomized controlled trials in acute and chronic myeloid leukemia, and myelodysplastic syndromes as governed by the CONSORT statement. *Ann. Epidemiol.* 19, 494–500.

2. Evaluation of the quality of the reporting of phase II clinical trials in oncology : A systematic review

2.1. Introduction

L'essai de phase II est une étape cruciale lors du processus de développement d'un nouveau médicament : Seules les molécules ayant prouvé une efficacité minimale au décours de cette phase, justifieront la poursuite de leur développement dans des essais comparatifs de phase III de plus grande échelle et qui nécessitent un financement beaucoup plus important.

Actuellement, en oncologie médicale, les molécules candidates à un développement en essai de phase III et celles potentiellement éligibles à une demande d'autorisation de mise sur le marché, se multiplient, notamment depuis l'avènement de nouvelles classes thérapeutiques telles que les thérapies moléculaires ciblées et l'immunothérapie. En parallèle, l'oncologie est la spécialité où le taux d'attrition est le plus élevé : Ainsi, le développement de nouvelles molécules en cancérologie est très souvent interrompu, et plus particulièrement au décours de l'essai de phase III ou le taux d'attrition moyen est de 11.3 à 14% supérieur par rapport aux autres spécialités médicales. Les essais de phase III sont des essais coûteux et qui peuvent poser un éventuel problème de perte de chance pour le patient recevant le traitement expérimental ne prouvant pas, in fine, sa supériorité par rapport au traitement de référence. A partir de ces observations, il paraît donc nécessaire d'élaborer des procédures standardisées et plus strictes d'évaluation de la méthodologie et du reporting des articles en amont du développement de la molécule en phase III.

Contrairement au reporting des essais cliniques randomisés de phase III, contrôlé par le biais de la déclaration CONSORT, les essais de phase II ne font pas l'objet actuellement de recommandation de reporting spécifique. L'objectif de cette revue de littérature a consisté à réaliser une synthèse de l'ensemble des publications ayant évalué la qualité du reporting des essais de phase II en oncologie et à identifier les outils d'évaluation utilisés pour cela.

2.2. Méthodes

Dans la présente revue systématique de la littérature, les critères d'inclusion pour les publications étaient : Articles de type revue systématique de la littérature, lettre à l'éditeur, dont l'objectif était d'évaluer la qualité du reporting et/ou la qualité méthodologique des essais de phase II en oncologie. Les publications écrites en anglais ou en français ont été incluses. Il n'y avait pas de limitation de date de parution pour les publications sélectionnées. La recherche a été réalisée via les bases de données MEDLINE et COCHRANE par un auteur (Julien Langrand Escure). La sélection des articles a été réalisée premièrement sur titre, puis dans une deuxième étape sur abstract puis sur texte intégral si la publication semblait en adéquation avec le sujet de la revue de littérature. Enfin, une recherche manuelle a également été réalisée dans la littérature grise, issue des données bibliographiques contenues des publications retenues au décours de la recherche sur bases de données.

Pour l'analyse des articles, une grille d'extraction de données a été élaborée par une équipe pluridisciplinaire, composée d'un méthodologiste (Aurélien Bourmaud), un statisticien (Fabien Tinquaut) et de cliniciens (Romain Rivoirard et Julien Langrand Escure). La grille a

dans un premier temps été testée et analysée, sur les trois premières publications sélectionnées. Les données ont été extraites par un auteur (Julien Langrand Escure). Les données recueillies pour l'analyse de la littérature étaient : L'année de publication de l'étude, le nom du journal ayant publié l'étude, le type de publication, la localisation tumorale étudiée, le nombre d'articles analysés, la période de sélection des articles, les différents items de reporting étudiés, les scores ou la fréquence de reporting des items étudiés, les facteurs associés à la variation de la qualité de reporting de l'item étudié et les recommandations potentielles et / ou discussions méthodologiques associées à la revue de la littérature.

3-3 Résultats

Les recherches réalisées dans les bases de données MEDLINE et COCHRANE avaient permis d'identifier 220 publications. Après sélection sur titre, sur abstract, sur texte intégral et après recherche dans la littérature grise, 13 articles ont finalement été retenus pour l'analyse.

Sur les 13 publications, 9 étaient publiées entre 2000 et 2009. La plupart des articles (10 articles sur 13) ont analysé le reporting des études de phase II d'oncologie quel que soit la localisation tumorale. 12 articles étaient des revues de littérature, 1 article était une lettre à l'éditeur.

Deux publications sur les 13 sélectionnées avaient évalué la qualité du reporting des essais de phase II à l'aide d'un score de reporting comprenant une checklist de plusieurs items. La première publication de 2014 de Grellety et al [193] avait porté sur la qualité de reporting de 156 articles d'études de phase II publiées en 2011. Les auteurs avaient élaboré un score global de qualité de reporting (Overall Quality Score ou OQS), comprenant 44 items à vérifier dans ces articles d'essais de phase II. Chaque item compris dans la liste couvrait une partie spécifique de l'article (Titre, introduction, méthodes, résultats, discussion) et comptait pour un point s'il était correctement décrit. Le score OQS médian était égal à 28 : $OQS_{\text{median}} = 28/44$ (min-max : 9-35). Les items les moins décrits étaient ceux en lien avec la méthodologie statistique et avec les objectifs de l'étude. Un second score avait également été développé par les auteurs, intitulé score KMS (Key Methodological Score). Ce score prenait en compte 3 items méthodologiques importants, comptant chacun pour un point lorsqu'il était présent dans l'article : La définition claire du critère de jugement principal, la justification claire du nombre de patients inclus dans l'étude et la définition à priori de la population dans laquelle a été évalué le critère de jugement principal et secondaire. Le score KMS médian était égal à 2/3 et 16.1% des études avaient un score KMS maximal de 3/3.

La deuxième publication de 2005 d'Ottaiano et al [194], a analysé la qualité de reporting de 141 articles d'études de phase I/II et de phase II en utilisant un index de qualité de reporting basé sur 11 items, avec un score global de qualité compris entre 0 et 100. Le score de qualité médian était égal à 53/100 (min-max : 18-86), avec une grande hétérogénéité entre les différents essais de phase II.

Les onze autres publications incluses pour analyse n'avaient pas utilisé de score global de qualité de reporting, mais elles s'étaient focalisées sur le reporting d'un ou de

plusieurs items précis. L'élément de reporting le plus étudié dans ces publications était le plan statistique de l'étude (63.6% des publications), suivi de la description du critère de jugement principal (27.3% des publications) et de la justification du calcul à priori du nombre de patients nécessaires (18.1% des publications). Le plan statistique des études de phase II était correctement rapporté dans moins de 50% des cas, pour 6 publications ayant évalué cet item. Le critère de jugement principal a été considéré comme bien rapporté dans 10 à 100% des articles, selon les séries analysées, avec un impact significatif de la date de publication de l'étude. Le reporting des éléments justifiant le calcul du nombre de patients à inclure dans l'essai était retrouvé pour 53 à 61% des études analysées.

Quatre études avaient apprécié l'évolution de la qualité du reporting au fil du temps. Ottaiano et al [194], n'avaient pas mis en évidence d'amélioration de l'index de qualité entre 1998 et 2002. Par contre, Thezenas et al [195] ainsi que Lee JJ et al [29] avaient observé une amélioration de la qualité du reporting du design statistique des essais de phase II au cours du temps. De la même manière, Nickolich et al [196] avaient démontré une amélioration de la qualité du reporting du critère de jugement principal, entre 1986 et 2010.

3-4 Discussion

Cette revue de littérature souligne tout d'abord le manque de publications (seulement 13) qui se sont intéressées à la qualité de reporting des essais de phase II en oncologie. Trois scores de qualité de reporting ressortent de cette analyse : score OQS, score KMS [193] et score d'index de qualité [194]. En ce qui concerne la qualité de la méthodologie des essais de phase II, trois critères majeurs ont été retrouvés dans plusieurs publications : Le plan statistique, le critère de jugement principal et la justification du calcul du nombre de patients à inclure.

L'index de qualité de reporting, utilisé dans l'étude de Ottaiano et al [194], a été utilisé pour évaluer des études publiées il y a plus de 15ans. Ce score d'évaluation de reporting semble donc peu adapté pour évaluer la qualité d'essais de phase II plus récents, portant sur les thérapies moléculaires ciblées ou sur les immunothérapies, dans lesquels la méthodologie utilisée est complètement différente par rapport aux essais de chimiothérapie cytotoxique. De plus, l'index de qualité ne prend pas en compte un élément de reporting important : La justification et la description du calcul du nombre de sujets à inclure. Grellety et al [193] avaient proposé deux scores de reporting. Le score OQS est un score très complet, composé d'une liste de 44 items devant être vérifiés et respectés par les auteurs avant soumission de leurs manuscrits à un journal médical. Le score KMS semble plus facile d'utilisation et plus reproductible, il est composé de seulement trois items correspondant à des éléments méthodologiques majeurs à vérifier dans les articles publiés.

Contrairement aux analyses qualitatives de la qualité du reporting, l'utilisation d'un score permet une standardisation et une reproductibilité de l'évaluation par les auteurs et les lecteurs. Les analyses qualitatives ont tendance à se concentrer sur un critère précis de reporting : 10/11 publications ayant réalisé une analyse qualitative se sont concentrées sur un seul critère, qui correspondait le plus souvent au reporting du plan statistique de l'étude.

Quatre études ont analysé l'évolution de la qualité du reporting des essais de phase II au cours du temps, montrant pour la plupart une amélioration de cette qualité de reporting pour les publications les plus récentes. Ceci pourrait s'expliquer par une attention et une application plus importante des auteurs, lors de l'étape de soumission de leur manuscrit, depuis la publication des recommandations CONSORT en 1996.

Les résultats de la revue de littérature mettent en lumière une qualité insuffisante du reporting des essais de phase II en oncologie : Grellety et al [193] ont décrit dans leurs résultats que seulement 16.1% des études avaient un score KMS maximal de 3/3, Ottaiano et al [194] avaient détaillé dans leur publication que le score de qualité médian était égal à 53/100 et plusieurs autres études avaient objectivé un reporting correct du plan statistique dans moins de 50% des essais de phase II.

Cette revue de littérature présente plusieurs limitations. Premièrement, il peut se poser le problème de la pertinence de l'évaluation de la qualité du reporting des essais thérapeutiques en général, en tant que critère de substitution pour l'évaluation de la qualité méthodologique de ces études. En effet, Soares et al [197] avaient déjà remarqué qu'une mauvaise qualité du reporting dans un essai clinique randomisé n'était pas systématiquement associée à une mauvaise méthodologie du protocole de l'essai thérapeutique. De plus, la qualité du reporting ne représente pas le seul facteur explicatif du taux d'échec élevé, observé pendant le développement en phase III des nouvelles molécules en oncologie : Certains facteurs méthodologiques semblent également jouer un rôle important, comme l'utilisation d'un critère de jugement principal inadéquat ou l'utilisation d'un design expérimental inapproprié afin d'évaluer la nouvelle molécule. D'autres facteurs sans lien avec la méthodologie pourraient aussi expliquer ce taux d'attrition élevé, comme : Le grand nombre de molécules testées dans des essais de phase II d'oncologie, l'index thérapeutique habituellement bas des traitements systémiques anticancéreux, les considérations économiques, les choix pris par les industriels, le manque de biomarqueur prédictif de réponse.

En conclusion, la qualité du reporting, tout comme la méthodologie, des essais de phase II en oncologie reste perfectible. Ce sujet est encore peu étudié et il n'existe actuellement pas de recommandation de reporting validée pour les essais de phase II. Il semble ainsi important d'actualiser les connaissances, sur la méthodologie et la qualité de reporting des essais de phase II en oncologie, en réalisant une revue systématique de la littérature, se focalisant sur des essais de phase II plus récents. L'évaluation du reporting pourrait être réalisée à l'aide de score reproductible comme le score KMS.

L'article original en texte intégral est disponible ci-dessous (en anglais).



Contents lists available at ScienceDirect

Critical Reviews in Oncology / Hematology

journal homepage: www.elsevier.com/locate/critrevonc

Evaluation of the quality of the reporting of phase II clinical trials in oncology: A systematic review



Romain Rivoirard^a, Julien Langrand-Escure^b, Mathieu Oriol^c, Fabien Tinquaut^c,
Franck Chauvin^{c,d,e}, Chloé Rancoule^b, Nicolas Magné^b, Aurélie Bourmaud^{c,d,*}

^a Medical Oncology Department, Lucien Neuwirth Cancer Institute, Saint Etienne, France

^b Radiotherapy Department, Lucien Neuwirth Cancer Institute, Saint Etienne, France

^c Hyg e Centre, Lucien Neuwirth Cancer Institute and Inserm, Clinical Investigation Center 1408, Saint-Etienne, France

^d Quality Safety Performance in Health (HESPER) EA7425, Lyon 1 University, Lyon, France

^e Jacques Lisfranc Medical School, Jean Monnet University, Saint Etienne, France

ARTICLE INFO

Keywords:
Reporting
Phase II
Clinical trials
Methodology
Oncology
Literature review

ABSTRACT

Objective: To describe the current state of knowledge concerning the quality of reporting in phase II clinical trials in oncology and to describe the various methods published allowing this quality evaluation.

Methods: databases including MEDLINE and COCHRANE were searched. Reviews and meta-analyses analyzing the quality of the reporting of phase II trials in oncology were included. Descriptive analysis of the results was performed.

Results: Thirteen publications were retained. Only 2 publications adopted a systematic approach of evaluation of the quality of reporting by overall scores. The Key Methodological Score (KMS), proposed by Grellety et al., gathering 3 items, seemed adapted for such an evaluation. A score of 3/3 was found in 16.1% of the 156 phase II trials analysed by this score. The other reviews used a qualitative analysis to evaluate the reporting, via an analysis of a single criterion, generally the statistical plan of the study. This item was considered as having been correctly reported in less than 50% of the analysed articles.

Conclusion: The quality of reporting in phase II trials in oncology is a field that has been investigated very little (13 publications). When it is studied, the estimated level of quality is not satisfactory, whatever the method employed. The use of an overall score of evaluation is a path which should be pursued, in order to get reliable results. It also seems necessary to propose strong recommendations, which would create a consensus for the methodology and the reporting of these studies.

1. Introduction

In oncology, phase II clinical trials play a key function in the development of new treatments and/or new therapeutic strategies. They aim at investigating the effectiveness and the toxicity of an experimental treatment or a combination of treatments. The objective is to select the molecules justifying the realization of a phase III clinical trial, thus ensuring the role of a “filter”. These tests are carried out on a very selected population of cancer patients.

Currently, in the field of oncology, the treatments that are candidates for phase III trials and those which are candidates for the obtention of an authorization to be marketed, are multiplying, particularly since the advent of targeted therapies. Parallel to this, oncology proves to be the specialty which has the highest attrition: the highest number of molecules whose development is interrupted, taking all the stages of

the drug development into consideration (Kola and Landis, 2004). At the stage of phase III trials in particular, oncology shows an average rate of attrition superior by 11.3%–4% in comparison to the average rate of attrition known in other specialties (Kola and Landis, 2004; DiMasi and Grabowski, 2007). These observations imply that the selection of experimental treatments under development in phase II trials, is of lower quality in the field of oncology, compared to that in other specialties. Phase III trials are more expensive and may cause a potential loss of chance for the participating patients when they prove to be negative. This observation, concerning the rate of attrition of the molecules in oncology, which is higher in phase III, must lead to the creation of procedures for the evaluation of the quality of the previous stages of the therapeutic development: phase II trials.

Contrary to the controlled randomized clinical trials, whose drafting is supervised by the CONSORT recommendations which are regularly

* Corresponding author at: Hyg e Centre, Lucien Neuwirth Cancer Institut, 108 bis avenue A. Raimond, 42 270 Saint Priest en Jarez, France.
E-mail address: aurelie.bourmaud@icloire.fr (A. Bourmaud).

<https://doi.org/10.1016/j.critrevonc.2018.02.014>

Received 19 July 2017; Received in revised form 4 December 2017; Accepted 26 February 2018
1040-8428/  2018 Elsevier B.V. All rights reserved.

updated (Begg et al., 1996; Moher et al., 2001; Schulz et al., 2010), in the case of clinical phase II trials in oncology, there is no consensus of precise and specific recommendations for the reporting and drafting of these trials. This is also the case when it comes to having a validated score of evaluation for methodological quality. However, Ivanova et al. (2016), showed that the phase II trials in oncology, which described the statistical design of the study in an incomplete way, presented a significantly higher probability of concluding positively on the effectiveness of the studied molecule. Whereas many other issues (as the large number of drugs tested, the historic low therapeutic index of cytotoxic agents, the costs related to treatment development or the lack of biomarker strategy to support patient stratification) may explain the high attrition rate in oncology trials, this tends to suggest that the quality of the reporting in the publications of phase II trials in oncology could have an impact on the future of the development of the experimental treatment.

The assumption that the strong rate of attrition in the clinical phase III trials in oncology can be, at least partly, explained by the methodology of the phase II trials, led us to carry out a systematic literature review, with the objectives of making an inventory of the publications having analysed the quality of the reporting of the clinical phase II trials in oncology, and of describing the various methods available for a reproducible systematic evaluation of the quality of the reporting in these studies.

2. Methods

2.1. Selection of the articles

The criteria of inclusion of the studies in this review of literature were: publications in the form of systematic reviews of phase II trials in oncology or leading articles or letters to the editor, whose objective were to evaluate the quality of the reporting of methodological criteria. Publications written in both French and English, were retained. There was no limited date for publication selected, with a last electronic request carried out in December 2015. Original articles, as well as the publications which related to a field other than oncology and recommendations, were excluded from the selection.

Research was carried out using the MEDLINE and COCHRANE databases, in order to identify the publications whose subject was related to the analysis of the reporting of the phase II trials in oncology, as well as the different items that are essential for reporting of good quality. The keywords used were “reporting” and “phase II”; the terms “neoplasms” and “Clinical Trials; Phase II as Topic/standards” were used as MeSH terms. The request formulated in MEDLINE was built in the following way: (“Clinical Trials; Phase II as Topic/standards” [MeSH Terms] AND “neoplasms” [MeSH Terms]) OR (“reporting” AND “phase II” AND “neoplasms” [MeSH Terms]). The request formulated in the COCHRANE database was: (“Phase II trial” AND “reporting” AND “neoplasm”); with a filter allowing us to retain only the studies of methodology (“methods studies”).

The selection of the studies was carried out initially on the basis of the titles of articles, highlighted by the request carried out in the MEDLINE and COCHRANE databases. Secondly, a reading of the summaries of the studies selected on their title, followed by a complete reading of the publication if the summary were in adequacy with the objective of our study, were carried out, leading to a final selection. Finally, a manual research in the grey literature was also carried out, using the bibliographical data resulting from the publications retained by electronic request.

2.2. Analysis of the articles

We were able to consult all of the publications retained for analysis, in their full version. A grid for the extraction of the data was built by a multidisciplinary team, composed of a methodologist, a statistician and

clinicians. This grid of extraction was then tested and modified, after having analysed 3 of the selected publications (Table S1). The variables were extracted by one of the authors. The data collected for the analyses of the literature which concerned the evaluation of the quality of the reporting were: the year of publication of the study, the name of the journal, the type of publication, the cancer site concerned, the number of analysed articles, the period of selection of the articles, the different items studied, scores or frequency of reporting of the studied items, the factors associated with the variation in quality of the reporting of the studied item and the potential recommendations and/or methodological discussions associated with the review of the literature.

No statistical analysis was carried out in this review of literature. The delivered results were purely descriptive and were expressed as a percentage.

3. Results

3.1. Selection of the articles

The request carried out in the MEDLINE database made it possible to obtain a result of 220 publications. Only 9 of these articles corresponded to our selection criteria and were conserved for analysis. The request carried out in the COCHRANE database did not highlight any article answering our selection criteria. In parallel, manual research in the grey literature made it possible to select 4 publications of interest. In total, 13 articles were finally selected (Lee, 2005; Nickolich et al., 2014; Kramar et al., 1996; Medioni et al., 2000; Perrone et al., 2003; Thezenas et al., 2004; Mariani and Marubini, 2000; Ottaiano et al., 2005; Walter and Krzyzanowska, 2012; Vickers et al., 2007; Riechelmann et al., 2008; El-Maraghi and Eisenhauer, 2008; Grellety et al., 2014). The selection process of the publications of interest is detailed in the flow-chart (Fig. 1).

3.2. Characteristics of the systematic review of literature

Four out of thirteen of the selected publications were published in only one Medical Journal: Journal of Clinical Oncology. These publications were most of the time published over 10 years between 2000 and 2009 (9 out of 13 publications). The oldest has been published in 1996 (Kramar et al., 1996) and the latest have been published in 2014 (Nickolich et al., 2014; Grellety et al., 2014). Most of those reviews (10 out of 13) assessed studies no matter of the cancer site. All of them were reviews but one (Riechelmann et al., 2008) which was a letter. The characteristics of the selected articles is summarized in Table S2 (online only Supplementary material).

3.3. Analysis

3.3.1. Evaluation of the reporting by a score

We focused on studies using a score to assess the quality of reporting, because scores are supposed to be more reliable, valid, discriminant than qualitative assessment. What's more they are standardized: they eliminate the variation of measurement, allowing generalization and comparison of results (Kline, 1986). Two of the 13 publications selected (15.4%), studied the quality of reporting of one or more criteria with an overall score: Grellety et al. (2014) and Ottaiano et al. (2005). Grellety et al. (2014) developed 2 scores of evaluation. The first score, named Overall Quality Score (OQS), gathers 44 items, each one counting for a point and covering the various sections of a scientific article (Title, Introduction, Methods, Results, Discussion). The section “Methods” was the most detailed, with items that showed particular interest in the aims of the study, in the selection of the patients, the description of the studied treatment, the methods of measurement of the criteria of judgement and with statistical methodology.

In the systematic review carried out by the authors, taking into account 156 tests of phase II trials of oncology published in 2011, the

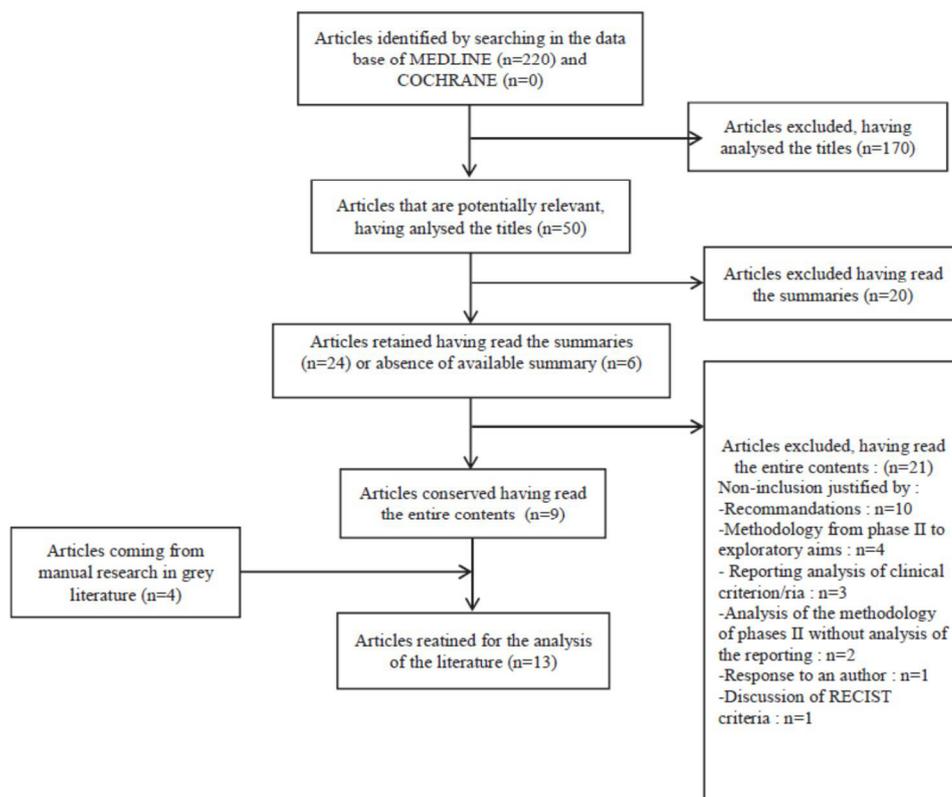


Fig. 1. Flow-chart showing the selection process used to retain publications in this systematic review.

median OQS was of 28/44 (range: 9–35). The items with the least information were those related to the sections treating statistical methodology and objectives of the study. The second score of quality developed by the authors, was entitled Key Methodological Score (KMS), and took into account 3 items, each one also counting for a point: the clear definition of a criterion of principal judgement, the clear justification of the number of patients included and the definition, a priori, of the population on which the principal and secondary judgement criteria were evaluated. The median KMS was estimated to 2/3. Only 16.1% of the analysed studies had a KMS equal to 3.

Ottaiano et al. (2005), reported an evaluation of the quality of reporting of 141 phase II or I/II trials in biotherapy or immunotherapy, using an index of quality of 11 items rated from 0 to 100. The conclusion was in favour of a great heterogeneity in the different phase II trials, with a median score of 53/100 (ranging between 18 and 86). The principal result concerning the evaluation of the reporting by an overall score, are summarized in Table 1.

3.3.2. Qualitative evaluation of the reporting

3.3.2.1. *Description by criteria.* The 11 other publications retained for analysis were not based on a systematic evaluation through a score, but on the evaluation of the frequency of the reporting of one or more criteria given (Table 2). Among them, the evaluation of the reporting of the statistical plan of the study was the criterion that was most frequently explored (63.6% of the publications) (Lee, 2005; Kramar et al., 1996; Medioni et al., 2000; Perrone et al., 2003; Thezenas et al., 2004; Mariani and Marubini, 2000; Walter and Krzyzanowska, 2012), followed by the criterion of principal judgement criterion (27.3% of the publications) (Nickolich et al., 2014; Walter and Krzyzanowska, 2012;

Riechelmann et al., 2008), the justification of the calculation of the sample size (18.1% of the publications) (Walter and Krzyzanowska, 2012; El-Maraghi and Eisenhauer, 2008), the correlation between the sample size, defined a priori, and the size of the sample that was reached in reality (9% of the publications) (Walter and Krzyzanowska, 2012) and the origin of the historical data of reference (9% of the publications) (Vickers et al., 2007).

The statistical plan of the study was reported in less than 50% of the cases for 6 of the 7 publications having evaluated this item (Lee, 2005; Kramar et al., 1996; Medioni et al., 2000; Perrone et al., 2003; Thezenas et al., 2004; Mariani and Marubini, 2000). A significant improvement of the quality of the reporting of the statistical plan in the phase II trials, was found to be associated with the period of publication, with the multicentric character, with the study of a single molecule, with the impact factor (IF) of the journal of publication, with the trial results and with a shorter period between initiation of the study and publication (Lee, 2005; Perrone et al., 2003; Thezenas et al., 2004; Mariani and Marubini, 2000). The principal judgement criterion was reported in 10 to 100% of the publications, depending on the series that were being analysed, with a significant impact of the period of publication of the study (Nickolich et al., 2014). The reporting of the elements justifying the sample size for the number of patients that were necessary, went from 53 to 61% in the analysed tests (Walter and Krzyzanowska, 2012; El-Maraghi and Eisenhauer, 2008) (Table 1).

3.3.2.2. *Evolution over time.* Four analyses studied the evolution of the quality of the reporting in the course of time (Table 1). Ottaiano et al. (2005), did not find significant improvement of the index of quality between 1998 and 2002. On the contrary, Thezenas et al. (2004) and

Table 1
Characteristics and main results of publications, having analysed the quality of reporting in phase II trials in oncology.

Article	Type of cancer studied	No of studies analysed	Period of selection of studies	Criteria studied	Median Score	Percentage of reporting of criteria studied	Factor(s) associated with the variation in the quality of reporting of the criteria studied
Lee (2005)	All	266	1986–2002	Evolution in the reporting of the design statistic, through time, in randomized phase II trials		46%	Increase in the reporting of the design statistic through time ($p < 0.001$)
Nicolich et al. (2014)	Lung cancers and cancers of small cell	38	1986–2010	Evolution in the reporting of principal judgement criteria through time (3 strata: 1986–1996, 1997–2005, 2006–2010)		1986–1996: 10% 1997–2005: 55,6% 2006–2010: 100%	Increase in the reporting of principal judgement criteria through time ($p = 0.01$)
Kramar et al. (1996)	All	83	1995	Reporting of the design statistic		28,8%	
Medioni et al. (2000)	All	124	1995–1998	Reporting of the design statistic		28%	
Perrone et al. (2003)	Breast cancer	145	1995–1999	Reporting of the design statistic		35,2%	Increased reporting of the design statistic, associated with: – Multicentre characteristic of the trial (OR = 3,24; IC95% [1,47–7,15]) – Study of a unique drug (OR = 2,35; IC95% [1,01–5,51]) – Publication in a journal with a higher IF ($p = 0.01$) – Shorter lapse of time between the beginning of the study and its publication ($p = 0,01$) – Increase in the reporting of the design statistic over the period of 5 years ($p < 0,001$)
Thezenas et al. (2004)	All	393 – 1995: 85 – 2000: 208	1995 et 2000	Evolution in the reporting of the design statistic between 1995 et 2000		1995: 15% 2000: 46%	Increase in the reporting of the design statistic associated with: – Result of the study was more frequently negative – IF of journal – Index of Quality < versus $\geq 50/100$ associated with the IF of the journal
Mariani and Marubini (2000)	All	295	1997	Reporting of the design statistic		19,7%	
Ottaviano et al. (2005)	Solid tumours	141	1998–2002	Index of Quality	53/100 (range 18–86)		
Walter and Krzyzanowska (2012)	Neuro-endocrine tumours	46 (39 phase II trials + 7 phase III trials)	2000–2011	Reporting of – Design statistic – Principal judgement criteria – Size of sample defined beforehand – Size of sample defined beforehand and that is reached Reporting of the origin of historical data		74% 72% 61% yes: 50% no: 11%, NA: 39% 13%	
Vickers et al. (2007)	All	70	2002–2005	Reporting of the origin of historical data			Historical data not mentioned or mentioned without respecting recommendations given by authors, associated with a result of the trial that is more often positive
Reichmann et al. (2008)	All	125	2005	Reporting of principal judgement criteria		73%	
ElMaraghi and Eisenhauer (2008)	Cancers of the breast, colon, lung, ovary, prostate and kidney with clear cells	51	NA–2006	Reporting of the justification of the calculation for the sample size among trials whose principal judgement criteria was the objective tumoral response OQS		53%	
Grealey et al. (2014)	All	156	2011	KMS: 1. clear definition of a principal judgement criterion 2. clear justification of number of patients included 3. definition à priori of the population on which the principal and secondary judgement criteria are evaluated	28/44 (range 9–35) 2/3	KMS: 1. 68,6% 2. 77,6% 3. 33,3%	Higher OQS associated with: – Recorded on clinicaltrials.gov – IF > 10 Higher KMS: No found factor of association

IF: Impact Factor; OQS: Overall Quality Score; KMS: Key Methodological Score.

Table 2
Frequency and type of criteria studied for the quality of their reporting in journals on phase II trials in oncology (n = 13).

Criteria studied	Number of publications (%)
Statistic Plan	7 (53.8)
Criteria of judgement	3 (23.1)
Justification of the sample size	2 (15.4)
Global Score	2 (15.4)
Origin of historical data	1 (7.7)
Size of planned sample reached	1 (7.7)

Lee (2005), brought back an improvement of the reporting of the statistical design of the phase II trials over time, respectively from 1995 to 2000 and 1986 to 2002. In the same way, Nickolich et al. (2014) showed an improvement of the quality of the reporting of the criterion of principal judgement between 1986 and 2010, in a series of phase II trials for small cell lung cancers.

4. Discussion

This systematic literature review underlines that very few publications (only 13) are concerned with investigating the quality of the reporting in oncology phase II trials. The methodological and editorial quality of the phase II trials in oncology thus seems to be a field that is rarely studied, in spite of the central role that these trials play in the development of future therapeutic strategies. Our meta-review highlights 3 scores (OQS and KMS (Grellety et al., 2014) and index of quality (Ottaiano et al., 2005)) and 3 key criterions (the statistical plan, the principal judgement criterion and the justification of the sample size calculation) for evaluation of the methodology of phase II oncology clinical trials which are essential to the description of these studies, and yet which are overlooked, in spite of multiple existing recommendations (Seymour et al., 2010; Brown et al., 2011; Booth et al., 2008; Chang et al., 2005).

Only 3 scores, making it possible to evaluate the quality of the reporting, were thus reported through 2 publications (Ottaiano et al., 2005; Grellety et al., 2014). The principal limits of the index of quality created by Ottaiano et al. (2005), are that it has become outdated (publication in 2004), especially as the methodology of phase II trials has evolved significantly since then, as well as the fact that it was developed for trials in biotherapy or immunotherapy. It is thus, in consideration of this last point of view, not generalizable with all trials for anti-cancer treatments. Moreover, this index of quality does not take into account a major criterion in the construction of a therapeutic trial: the justification and the description of the calculation of the sample size. Grellety et al. (2014), proposed 2 scores. The OQS is an exhaustive score, with 44 items described by the authors, and which may be considered as a “checklist” to be respected by those undertaking and writing about future phase II trials in oncology. As regards the evaluation of the quality of reporting a study, the use of the KMS seems easier, since it consists of only 3 items, quickly making it reproducible for the reader. These 3 items are indisputably key methodological elements for the drafting of phase II trials.

Contrary to qualitative analyses of the reporting, the use of score allows a standardization and a reproducibility of the evaluation for the authors and/or readers. The evaluation of the quality of the reporting through a score also allows a comprehensive assessment of the methodology of the publication. Indeed, the qualitative analyses often tend to be focused on a precise criterion: 10 of the 11 publications carrying out a qualitative analysis of the reporting targeted only one criterion, this generally being the reporting of the statistical plan of the study (Lee, 2005; Nickolich et al., 2014; Kramar et al., 1996; Medioni et al., 2000; Perrone et al., 2003; Thezenas et al., 2004; Mariani and Marubini, 2000; Vickers et al., 2007; Riechelmann et al., 2008; El-Maraghi and Eisenhauer, 2008). Only Walter and Krzyzanowska

(2012), studied the reporting of various items in the same review of literature of 39 phase II trials and 7 phase III trials in a qualitative way.

It is interesting to notice that few studies were interested in exploring the evolution of the quality of the reporting according to time. Some analyses of the literature were carried out in a transverse way, considering one single year of publication (Kramar et al., 1996; Mariani and Marubini, 2000; Riechelmann et al., 2008; Grellety et al., 2014). Only 4 analyses studied the evolution of the quality of the reporting in the course of time showing a trend in the improvement of the quality of the reporting over time. This could be explained by the fact that the authors had become more aware of the great principles concerning the drafting of therapeutic trials, since the publication of the recommendations CONSORT in 1996 (Begg et al., 1996).

The results of our systematic review of the literature highlight a heterogeneous, yet an overall unsatisfactory quality of the reporting in phase II trials, and in particular, showing a median KMS of 2/3 and a score of 3/3 in only 16.1% of the tests analysed by Grellety et al. (2014), a median index of quality of 53/100 within the studies analysed by Ottaiano et al. (2005) and a reporting of the statistical plan of the study that is under 50% for the majority of the publications having analysed this criterion (Lee, 2005; Kramar et al., 1996; Medioni et al., 2000; Perrone et al., 2003; Thezenas et al., 2004; Mariani and Marubini, 2000). These results are coherent with those of an analysis of the literature by Ottaiano et al. (2006). The authors analysed a series of 80 therapeutic tests published in 2005. They showed that the phase II trials (n = 39) were published in journals with a lower IF compared to phase III trials (n = 41). The authors explained this observation by giving more weight to a defect in the methodological quality of the phase II trials, than to the editorial policy of the various journals.

Our review of the literature presents several limits. Firstly, we need to discuss the relevance of the analysis of the quality of the reporting in the therapeutic trials in general, as a criterion of substitution for the evaluation of the methodological quality of these studies. Indeed, Soares et al. (2004) showed that poor quality in the reporting of randomized therapeutic phase III trials, was not systematically associated with bad methodology of the protocol of the therapeutic trial. This observation could easily be found in the phase II trials. We may also discuss the real impact of the quality of the reporting and the methodological quality of the phase II trials on the excess of experimental treatments leading up to phase III, in spite of the results reported by Ivanova et al. (2016). Others hypotheses which can be put forward to explain the high rate of failure in phase III of the molecules in oncology, are either the use of inadequate principal judgement criteria, which are not representative of a significant clinical benefit in routine, at the stage of phase II; or the use of inappropriate experimental designs to assess the drugs. To answer that specific question, it would be necessary to analyze previous studies performed with innovative designs such as Bayesian or adaptive design. Also, non-methodologic factors could also explain this high attrition rate as the large number of drugs tested in oncology phase II trials, the historic low therapeutic index of cytotoxic agents leading to excess of toxicity or insufficient efficacy, economic considerations, industrial choices and/or strategy or the lack of biomarker strategy to support patient stratification. Next, our conclusions must be carried out with caution since they relied on only 12 reviews, and only 2 reviews using a score. What's more, the 3 scores used and applied to assess quality of reporting in these 2 reviews have not been further validated, so their internal validity can be interrogate. Another limit of this study is that those reviews are now becoming rather old (the latest was published in 2011) and certainly failed to capture the strikingly changing epidemiology of phase 2 trials we are now witnessing, as regards to methodologies and implementation. This information per se (the lack of recent data about reporting of phase 2 trials) is a plea to investigate further phase 2 trials new methodologies reporting. What's more although the research carried out in the literature intends to be exhaustive, we cannot exclude the possibility of the non-taking into account of publications which could have backed up

the results presented. Finally, a limitation concerning methodological construction of our study is that only one author has folded the data extraction grid. A double-checking by a second author of the 13 selected articles could have been performed to improve the quality of data collection. Moreover, the data extraction grid was only tested on 3 articles, which is a small sample, but is justified by the small number of articles selected for analysis in this meta-review.

In conclusion, because the experimental treatments play the role of a “filter”, phase II clinical trials are one of the cornerstones of the development of anti-cancer treatments. The methodological rigor of a clinical trial is an absolute prerequisite for the further development of the reported results. To evaluate this quality, the only tool available to the scientist is the reading of the report of the scientific experiment. This report must thus be factual and exhaustive and must allow the reproducibility of the experiment. Today it is clear that the quality of the reporting, as well as that of the methodology of phase II trials in oncology, remain perfectible and debated. Little data exists in the recent literature on this subject. It is necessary to update our knowledge as concerns the methodological quality of these studies, for example through an exhaustive review of literature of phase II trials in oncology, published in recent years, or by investigating the reporting of new phase 2 trials designs and mode of analysis, as well as giving preference to the use of a systematic analysis via a KMS score that has already been published. These results are a plea for the drafting of recommendations on the methodological construction of these studies and on their reporting, like those we see within the framework of other types of studies.

Conflict of interest

None.

Funding

No funding.

Contributions

RR, AB, NM and FC conceived the project. JLE and MO collected all the data. JLE, MO and FT did the statistical analyses. JLE wrote the first draft and revised drafts of the manuscript.

RR, JLE, MO, FT, FC, CR, NM and AB critically revised the manuscript for important intellectual content. The final version of the manuscript was approved by all the authors.

AB, JLE and RR had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. AB is the guarantor for the study.

Acknowledgment

The authors would like to thank Susan Guillaumond (Activenglish) for translation and editing services.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.critrevonc.2018.02.014>.

References

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al., 1996. Improving

- the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 276 (August (8)), 637–639.
- Booth, C.M., Calvert, A.H., Giaccone, G., Lobbezoo, M.W., Eisenhauer, E.A., Seymour, L.K., 2008. Design and conduct of phase II studies of targeted anticancer therapy: recommendations from the task force on methodology for the development of innovative cancer therapies (MDICT). *Eur. J. Cancer* 44 (January (1)), 25–29.
- Brown, S.R., Gregory, W.M., Twelves, C.J., Buyse, M., Collinson, F., Parmar, M., et al., 2011. Designing phase II trials in cancer: a systematic review and guidance. *Br. J. Cancer* 105 (July (2)), 194–199.
- Chang, S.M., Reynolds, S.L., Butowski, N., Lamborn, K.R., Buckner, J.C., Kaplan, R.S., et al., 2005. GNOSIS: guidelines for neuro-oncology: standards for investigational studies—reporting of phase 1 and phase 2 clinical trials. *Neuro-Oncology* 7 (4), 425–434.
- DiMasi, J.A., Grabowski, H.G., 2007. Economics of new oncology drug development. *J. Clin. Oncol.* 25 (January (2)), 209–216.
- El-Maraghi, R.H., Eisenhauer, E.A., 2008. Review of phase II trial designs used in studies of molecular targeted agents: outcomes and predictors of success in phase III. *J. Clin. Oncol.* 26 (March (8)), 1346–1354.
- Grellety, T., Petit-Moneger, A., Diallo, A., Mathoulin-Pelissier, S., Italiano, A., 2014. Quality of reporting of phase II trials: a focus on highly ranked oncology journals. *Ann. Oncol.* 25 (February (2)), 536–541.
- Ivanova, A., Paul, B., Marchenko, O., Song, G., Patel, N., Moschos, S.J., 2016. Nine-year change in statistical design, profile, and success rates of Phase II oncology trials. *J. Biopharm. Stat.* 26 (1), 141–149.
- Kline, P., 1986. *A Handbook of Test Construction: Introduction to Psychometric Design*. Methuen, New York.
- Kola, I., Landis, J., 2004. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3 (8), 711–716.
- Kramar, A., Potvin, D., Hill, C., 1996. Multistage designs for phase II clinical trials: statistical issues in cancer research. *Br. J. Cancer* 74 (8), 1317.
- Lee, J.J., 2005. Randomized phase II. Designs in cancer clinical trials: current status and future directions. *J. Clin. Oncol.* 23 (July (19)), 4450–4457.
- Mariani, L., Marubini, E., 2000. Content and quality of currently published phase II cancer trials. *J. Clin. Oncol.* 18 (January (2)), 429–436.
- Medioni, J.R., Rycke, Y.D., Asselain, B., 2000. New trends in assessment in anticancer treatments by phase II clinical trials. *Bull. Cancer (Paris)* 87 (July (7–8)), 551–556.
- Moher, D., Schulz, K.F., Altman, D.G., 2001. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Med. Res. Methodol.* 1 (1), 2.
- Nickolich, M., Babakoochi, S., Fu, P., Dowlati, A., 2014. Clinical trial design in small cell lung cancer: surrogate end points and statistical evolution. *Clin. Lung Cancer* 15 (May (3)), 207–212.
- Ottaiano, A., Mollo, E., Di Lorenzo, G., Pisano, C., Di Maio, M., Barletta, E., et al., 2005. Prospective clinical trials of biotherapies in solid tumors: a 5-year survey. *Cancer Immunol. Immunother.* 54 (January (1)), 44–50.
- Ottaiano, A., Facchini, G., Nasti, G., Budillon, A., Caraglia, M., 2006. Phase II studies of anticancer chemotherapy: indirect evidence of poor quality. *Ann. Oncol.* 18 (November (2)) 403–403.
- Perrone, F., Di Maio, M., De Maio, E., Maione, P., Ottaiano, A., Pensabene, M., et al., 2003. Statistical design in phase II clinical trials and its application in breast cancer. *Lancet Oncol.* 4 (5), 305–311.
- Riechelmann, R., Dounaevskaia, V., Krzyzanowska, M., 2008. Quality of reporting primary outcomes in phase II cancer trials. *J. Clin. Oncol.* 26 (November (33)), 5486–5488.
- Schulz, K.F., Altman, D.G., Moher, D., for the CONSORT Group, 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med.* 7 (March (3)), e1000251.
- Seymour, L., Ivy, S.P., Sargent, D., Spriggs, D., Baker, L., Rubinstein, L., et al., 2010. The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the national cancer institute investigational drug steering committee. *Clin. Cancer Res.* 16 (March (6)), 1764–1769.
- Soares, H.P., Daniels, S., Kumar, A., Clarke, M., Scott, C., Swann, S., et al., 2004. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 328 (January (7430)), 22–24.
- Thezenas, S., Duffour, J., Culine, S., Kramar, A., 2004. Five-year change in statistical designs of phase II trials published in leading cancer journals. *Eur. J. Cancer* 40 (May (8)), 1244–1249.
- Vickers, A.J., Ballen, V., Scher, H.I., 2007. Setting the bar in phase II trials: the use of historical data for determining go/no go decision for definitive phase III testing. *Clin. Cancer Res.* 13 (February (3)), 972–976.
- Walter, T., Krzyzanowska, M.K., 2012. Quality of clinical trials in gastroenteropancreatic neuroendocrine tumours. *Neuroendocrinology* 96 (3), 238–248.

3. Outcomes definitions and statistical tests in oncology studies: A systematic review of the reporting consistency

3.1. Introduction

Au terme de la première étude décrite ci-dessus, il avait été observé plusieurs constatations : Premièrement, le reporting des études observationnelles en oncologie avaient été peu analysé (une seule revue de littérature sur les 58 revues incluses). Deuxièmement, la définition du critère de jugement principal dans les Essais cliniques Randomisés Contrôlés (ECR) était rapportée de façon imprécise pour la moitié des revues de littératures ayant étudié cet élément de reporting. Cependant, aucun résultat n'avait pu être trouvé concernant le reporting du critère de jugement principal des ECR d'oncologie qui n'étaient pas basés sur des données censurées. Il en était de même pour le reporting des variables d'intérêt des études observationnelles en oncologie. Troisièmement, le reporting du test statistique utilisé pour le critère principal des études en oncologie n'avait pas été spécifiquement analysé.

Une revue systématique de la littérature a donc été réalisée, dont l'objectif principal était d'évaluer la qualité du reporting du critère de jugement principal (pour les ECR) et des variables d'intérêt analysées (pour les études observationnelles), définis dans la partie méthodes des publications d'études en oncologie. L'objectif secondaire était d'analyser la concordance du reporting des tests statistiques, entre la partie méthodes et la partie résultats, utilisés pour répondre à l'objectif principal de ces études. Dans un second temps, une analyse complémentaire a été réalisée, afin d'identifier d'éventuelles co variables dans ces études, corrélées à une meilleure concordance du reporting des tests statistiques. La revue de littérature avait notamment pour but d'analyser le reporting des études observationnelles en oncologie, peu décrits jusqu'à présent dans la littérature.

2-2 Méthodes

Dans un premier temps, les auteurs ont procédé à une phase de sélection de journaux médicaux ayant publié des études d'oncologie. A partir d'une liste de 19 journaux sélectionnés de manière pluridisciplinaire, 3 groupes ont été formés : Un premier groupe de journaux dits « généralistes » dont l'Impact Factor (IF) était supérieur à 6, un deuxième groupe de journaux spécialisés en oncologie dont l'IF était supérieur à 6 et un troisième groupe de journaux d'oncologie dont l'IF était compris entre 4 et 6. Un auteur (Fabien Tinquaut) a ensuite conduit un tirage au sort, afin de sélectionner un journal dans chacun de ces groupes. Les trois journaux finalement retenus étaient le British Medical Journal (BMJ, groupe des journaux dits « généralistes »), Annals of Oncology (AoO, groupe des journaux d'oncologie à $IF > 6$) et le British journal of Cancer (BjC, groupe des journaux d'oncologie à $4 < IF \leq 6$). Les archives de ces trois journaux ont été explorées, afin d'identifier tous les articles originaux publiés entre Mars 2009 et Mars 2014.

Les articles d'études éligibles pour une inclusion dans la revue de littérature, devaient être des articles d'études interventionnelles ou observationnelles, écrits en anglais, se focalisant sur un sujet en lien avec l'oncologie, publiés entre Mars 2009 et Mars 2014 dans le

BMJ, AoO ou BjC. L'objectif principal de l'étude devait être accompagné d'un test statistique correspondant pour que l'article soit incluable pour analyse.

Les articles d'études basées sur des données censurées ou ayant la survie comme critère principal ont été exclus de l'analyse pour deux raisons : Premièrement, les tests statistiques utilisés dans ces études sont très spécifiques et peu variés. Deuxièmement, ces articles d'études sont le plus souvent publiés dans un contexte d'obtention d'autorisation de mise sur le marché ou d'extension d'indication et sont soigneusement analysées par les autorités sanitaires, ce qui dépasse le processus standard de publication des autres études. L'inclusion de telles études avait donc été considéré par les auteurs comme un biais potentiel.

A partir de l'ensemble des articles identifiés, une première étape de sélection a été réalisée, sur titre et sur abstract, de façon indépendante par deux auteurs (Romain Rivoirard et Vianney Duplay). Les articles potentiellement éligibles ont ensuite été sélectionnés sur texte intégral. L'ensemble des articles discordants entre les deux sélections réalisées ont été relus, puis inclus ou non, par un troisième auteur méthodologiste (Aurélie Bourmaud).

Pour chaque article, un auteur a procédé à l'extraction des données suivantes (Romain Rivoirard), ensuite vérifiées par un deuxième auteur (Mathieu Oriol) : Nom du journal, année de publication, nombre de patients inclus, localisation du cancer étudié, type d'étude (observationnelles ou interventionnelle), description du critère de jugement principal dans la partie méthodes (pour les études interventionnelles), définition des variables dans la partie méthodes (pour les études observationnelles), description dans la partie méthodes du type et du nom du test statistique principal utilisé pour le critère de jugement principal ou pour les variables, le type et le nom du test statistique utilisé pour l'évaluation du critère principal ou des variables dans la partie Résultats, significativité des résultats pour le critère principal, conclusion concordante avec le résultat du critère principal.

Deux items ont été recueillis, afin de décrire la qualité du reporting du critère principal. Le premier intitulé « description des variables » pour les études observationnelles, le second « description du critère de jugement principal » pour les études interventionnelles. Pour les publications d'études observationnelles, la déclaration STROBE a été utilisée, afin de pouvoir définir si l'item « description des variables » était rapporté dans les articles. Pour les publications d'études interventionnelles, la déclaration CONSORT a été utilisée pour apprécier la qualité du reporting de l'item « description du critère de jugement principal ».

Concernant la concordance des tests statistiques entre les parties méthodes et résultats, dans les articles sélectionnés, seulement un test par article a été considéré : Il s'agissait du test statistique utilisé pour évaluer le critère principal de chaque étude. Les publications d'études ont été par la suite catégorisées en trois groupes : Un premier groupe avec une concordance parfaite méthodes/résultats pour le test statistique principal, un second groupe dans lequel le nom et le type de test statistique principal était correctement rapporté dans méthodes mais ces données n'étaient pas retrouvées dans résultats, et un troisième groupe qui incluait tous les articles présentant une discordance évidente entre le test statistique principal décrit dans méthodes et celui utilisé dans résultats.

Afin d'identifier les facteurs potentiellement corrélés à une concordance parfaite du test statistique principal entre méthodes et résultats, une régression logistique a été réalisée. Tous les articles d'études ayant une concordance parfaite ont été cotés « 1 », tandis que ceux qui étaient inclus dans le deuxième ou le troisième groupe ont été cotés « 0 ». Les variables testées dans l'analyse univariée étaient : Nom du journal, année de publication, nombre de patients inclus, localisation de la tumeur, type d'étude (observationnelle ou interventionnelle), conclusion concordante avec le résultat principal, type de test, critère d'évaluation principal décrit dans méthodes. Les Odds Ratios (OR), leurs intervalles de confiance à 95% et leurs valeurs de p bilatérales ont été estimés par régression logistique conditionnelle. Une analyse multivariée a été réalisée avec toutes les co-variables disponibles, puisqu'il n'y en avait que 8, contre 826 articles à analyser. Les analyses ont été effectuées avec le logiciel R, version 3.1.2 (<http://www.R-project.org>). Le risque- α était fixé à 5%. Le niveau de signification a été défini pour les valeurs $p \leq 0,05$.

2-3 Résultats

Au terme de la phase de sélection, 826 articles correspondaient aux critères d'inclusion et d'exclusion et ont été inclus dans l'analyse. Il y avait 698 articles d'études observationnelles (84.5% des publications) et 128 articles d'études observationnelles (15.5% des publications), dont 29 essais de phase III (3.5% des publications). Lorsqu'un seul site tumoral était étudié, les localisations les plus fréquemment rapportées étaient : Le cancer du sein (21.8% des articles), le cancer colo-rectal (11.3% des articles) et les cancers gynécologiques (7% des articles).

Les variables étudiées étaient correctement rapportées dans la section méthodes pour tous les articles d'études observationnelles. Une association significative entre exposition et maladie avait été observée pour 618 articles d'études observationnelles (88.6% des publications). Le critère de jugement principal était clairement rapporté dans la section méthodes pour 109 études interventionnelles (85.2% des publications). Le résultat concernant ce critère de jugement principal était statistiquement significatif pour 87 articles d'études interventionnelles (68% des publications).

Pour les études observationnelles, le test rapporté pour l'analyse principale était paramétrique pour 88.8% des articles. Parmi ces tests, la régression logistique était le plus fréquemment utilisé (416 articles d'études ou 59.6% des cas). La concordance de reporting pour le test statistique principal, entre les parties méthodes et résultats, a été évaluée comme parfaite pour 295 articles (42.2% des publications) d'études observationnelles. Le nom du test statistique n'était pas rappelé dans 385 articles (55.2% des cas), tandis qu'une discordance pour le reporting de ce test statistique a été observé dans 18 articles d'études observationnelles (2.6% des cas).

Concernant les études interventionnelles, le test principal utilisé était essentiellement paramétrique (89 articles, 69.5% des cas). Le test le plus fréquemment utilisé était le test chi2 (21 études, 16.4% des cas). Une concordance parfaite de reporting du test statistique principal, entre les parties méthodes et résultats, a été observée pour 43 articles (33.6% des cas). Le nom du test statistique n'était pas rapporté dans la section résultats pour 80 articles (62.5% des

cas). Une discordance évidente de reporting du test statistique principal avait été noté pour 5 articles (3.9% des cas).

Lors de la réalisation de la régression logistique, 338 articles (40.9% des publications) ont été inclus dans le groupe côté « 1 », correspondant au groupe d'articles ayant une concordance parfaite de reporting du test statistique principal entre méthodes et résultats. Les 488 autres articles (59.1% des publications) ont été inclus dans le groupe côté « 0 ». Au décours de l'analyse univariée, 3 caractéristiques d'études semblaient associées à une concordance parfaite de reporting du test statistique principal : Le nombre de patients inclus, le nom du journal et le type d'étude (observationnelle/interventionnelle). À la suite de l'analyse multivariée, le nombre de patients inclus demeurait une variable indépendante, associée à la concordance : Odds ratio ajusté ou aOR pour le groupe 3 (nombre de patients inclus dans l'étude $>2.57 \cdot 10^3$ patients) comparé au groupe 1 (nombre de patients inclus dans l'étude <160 patients) aOR= 0.52 [0.31-0.89], $p=0.009$.

2-4 Discussion

Parmi les 826 articles inclus dans la revue systématique de littérature, une majorité était constituée de publications d'études observationnelles (698 articles ou 84.5%). Le reporting des variables et des tests statistiques dans la section méthodes a été considéré comme bon (présent dans 100% et 97.7% des publications, respectivement). Cependant, seulement 295 articles d'études observationnelles (42.2% des articles) avaient une concordance parfaite de reporting du test statistique principal entre la section méthode et résultats. Ceci a également été constaté pour les études interventionnelles, où cette concordance parfaite de reporting n'était retrouvée que pour 43 articles (33.6% des publications). Ceci peut s'expliquer en grande partie par le défaut de reporting du nom du test statistique principal dans la section résultats (situation observée pour 54.9% des publications d'études observationnelles et pour 62.5% des publications d'études interventionnelles). Dans cette situation la concordance ne peut donc pas être évaluée. En réalité, chaque journal établit et publie des recommandations spécifiques pour l'analyse des données issues des études et il n'existe pas d'attitude consensuelle pour cet aspect du peer review. Il est également important de noter que les déclarations de reporting, comme STROBE pour les études observationnelles et CONSORT pour les études interventionnelles de type ECR, ne contiennent pas d'item spécifique pour le rappel du nom du test statistique utilisé dans la section résultats. La reproductibilité de ces types d'études ne peut ainsi pas être assurée.

Les résultats de cette étude suggèrent qu'une vérification systématique de ces aspects statistiques du reporting doit être réalisée, lors du processus de reviewing, afin de limiter au maximum le manque de transparence. L'ajout, au sein des journaux médicaux, d'un relecteur compétent en statistiques, devrait donc être obligatoire.

Les résultats de l'analyse multivariée ont révélé que la variable « nombre de patients inclus dans l'étude » était un facteur indépendant, associé à une concordance parfaite de reporting du test statistique principal, entre la section méthodes et résultats. Une des explications plausibles est que, dans les études avec peu de patients inclus, les hypothèses paramétriques ne sont pas toujours applicables. Par conséquent, les auteurs sont

particulièrement attentifs au type de test statistique utilisé et reportent ces tests aussi bien dans la section méthodes que dans la section résultats.

La présente revue systématique de littérature présente plusieurs limitations. Premièrement, l'analyse n'a porté que sur trois journaux. Les résultats obtenus méritent donc d'être confirmés, même si les trois journaux sélectionnés sont assez représentatifs de la littérature en oncologie. Les résultats obtenus concernent essentiellement le reporting des articles d'études observationnelles (84.5% des publications), il n'est donc pas possible de conclure définitivement pour les études interventionnelles.

En conclusion, le reporting des variables pour les publications d'études observationnelles et le reporting du critère de jugement principal pour les publications d'études interventionnelles, en oncologie, est satisfaisant. Une discordance de reporting pour le test statistique principal utilisé, entre les parties méthodes et résultats, a été identifiée dans 23 articles. Ceci constitue un problème méthodologique important, et cet élément de reporting n'est actuellement pas pris en compte dans les déclarations CONSORT et STROBE. Les auteurs et le comité de relecture encouragent donc à vérifier la concordance de ces tests statistiques.

L'article original en texte intégral est disponible ci-dessous (en anglais).

RESEARCH ARTICLE

Outcomes Definitions and Statistical Tests in Oncology Studies: A Systematic Review of the Reporting Consistency

Romain Rivoirard^{1*}, Vianney Duplay^{2,3}, Mathieu Oriol^{2,3}, Fabien Tinquaut^{2,3}, Franck Chauvin^{2,3}, Nicolas Magne⁴, Aurelie Bourmaud^{2,3}

1 Department of Medical Oncology Lucien Neuwirth Cancer Institute, Saint Priest en Jarez, France, **2** Department of Public Health, Hyg e Center, Saint Priest en Jarez, France, **3** Inserm, Clinical Investigation Center 1408, 42055 Saint-Etienne, France, **4** Department of Radiation Therapy Lucien Neuwirth Cancer Institute, Saint Priest en Jarez, France

* Romain.RIVOIRARD@icloire.fr



CrossMark
click for updates

Abstract

Background

Quality of reporting for Randomized Clinical Trials (RCTs) in oncology was analyzed in several systematic reviews, but, in this setting, there is paucity of data for the outcomes definitions and consistency of reporting for statistical tests in RCTs and Observational Studies (OBS). The objective of this review was to describe those two reporting aspects, for OBS and RCTs in oncology.

Methods

From a list of 19 medical journals, three were retained for analysis, after a random selection: British Medical Journal (BMJ), Annals of Oncology (AoO) and British Journal of Cancer (BJC). All original articles published between March 2009 and March 2014 were screened. Only studies whose main outcome was accompanied by a corresponding statistical test were included in the analysis. Studies based on censored data were excluded. Primary outcome was to assess quality of reporting for description of primary outcome measure in RCTs and of variables of interest in OBS. A logistic regression was performed to identify covariates of studies potentially associated with concordance of tests between Methods and Results parts.

Results

826 studies were included in the review, and 698 were OBS. Variables were described in Methods section for all OBS studies and primary endpoint was clearly detailed in Methods section for 109 RCTs (85.2%). 295 OBS (42.2%) and 43 RCTs (33.6%) had perfect agreement for reported statistical test between Methods and Results parts. In multivariable analysis, variable "number of included patients in study" was associated with test consistency: aOR (adjusted Odds Ratio) for third group compared to first group was equal to: aOR Grp3 = 0.52 [0.31–0.89] (P value = 0.009).

OPEN ACCESS

Citation: Rivoirard R, Duplay V, Oriol M, Tinquaut F, Chauvin F, Magne N, et al. (2016) Outcomes Definitions and Statistical Tests in Oncology Studies: A Systematic Review of the Reporting Consistency. PLoS ONE 11(10): e0164275. doi:10.1371/journal.pone.0164275

Editor: William B. Coleman, University of North Carolina at Chapel Hill School of Medicine, UNITED STATES

Received: April 26, 2016

Accepted: September 22, 2016

Published: October 7, 2016

Copyright:  2016 Rivoirard 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

Variables in OBS and primary endpoint in RCTs are reported and described with a high frequency. However, statistical tests consistency between methods and Results sections of OBS is not always noted. Therefore, we encourage authors and peer reviewers to verify consistency of statistical tests in oncology studies.

Introduction

In oncology, quality and methodology of published clinical studies are essential to support an informed decision making [1]. Since 1996 and publication of the Consolidated Standards of Reporting of Trials (CONSORT) statement [2], many productive efforts have been made to improve the quality of reporting for randomised controlled trials [3]. Currently, it exists many other reporting guidelines for enhancing the quality of a variety of study types: For example, The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for observational studies [4], Preferred Reporting Items for systematic reviews and meta-Analyses (PRISMA) for systematic reviews and meta-analyses [5] and reporting recommendations for tumor mARKer prognostic studies (REMARK) for tumor marker studies [6]. The peer review system [7] is mandatory in almost all journals and improves the overall quality of reporting for publications [8].

Despite an increasing rate of publications analyzing the quality of reporting of studies in oncology [9], few systematic reviews have specifically analyzed the reporting of the outcomes definition. For interventional studies, the item relating to the description of pre-specified primary and secondary judgment criteria measures is an essential methodological point in the CONSORT checklist [10]. For observational studies, the item concerning the definition of variables (outcomes, exposures, predictors, potential confounders and potential effect modifiers) is a critical methodological criteria in the STROBE checklist [4]; clear descriptions, and steps taken to adhere to them are particularly important for any disease condition which are of primary interest in the study. Moreover, an element of reporting evaluation has not been studied in previous reviews: the item entitled "Statistical methods". This item enables to assess if articles provide a clear and exhaustive description of statistical methods, including statistical tests. But no recommendation exist, either in the CONSORT or in the STROBE, that would advise to check if statistical tests used in Results section are consistent with those described in Methods. Discrepancies between the statistical tests described in methods and tests really performed in results can lead to bias and, so, interfere with interpretation of the results.

Therefore, the main objective of the present systematic review was to describe quality of reporting for primary outcome measure in oncology clinical trials and for variables of interest in oncology observational studies. For the second objective, we investigated characteristics of manuscripts associated with a perfect consistency for statistical tests.

Methods

Search strategy

A list of medical journals has first been developed, selected on the following criteria: Non-organ specific journals, which publish articles dealing with cancer. This list was divided into 3 groups: A first group of generalist journals with Impact Factor (IF) above 6, a second group of journals specialized in oncology with high IF (above 6 IF) and a third group of oncology

journals with middle IF (above 4 IF) [11]. Selection of these journals was decided in a multidisciplinary way. The list of the 19 Journals was as follow: the first group made of Generalist journals: New England Journal of Medicine, Lancet, Journal of the American Medical Association, Annals of Internal Medicine, Plos Med, British Medical Journal, JAMA internal Medicine; the second group made of oncology journals with an Impact Factor above 6: Cancer Research, Journal of Clinical Oncology, Lancet Oncology, Journal of National Cancer Institut, CA: A Cancer Journal for Clinicians, Annals of Oncology; the third group made of oncology journals with an Impact factor above 4: British Journal of Cancer, European Journal of Cancer, The Oncologist, Cancer, International Journal of Cancer, Journal of the National Comprehensive Cancer Network. Among the 19 journals the statistician (F. Tinquaut) conducted a random selection: For each group a number between 1 and the number of Journals contained in each group was assigned to each journal. A random number was then drawn by group (following a discrete uniform distribution between 1 and the maximal number per group). The numbered corresponding Journal was then selected. Archives of the three included journals (British Medical Journal (BMJ), Annals of Oncology (AoO) and British Journal of Cancer (BjC)) were searched so that to identify all original articles, published between March 2009 and March 2014. The last search was performed in February 2015 [12].

Study selection

Observational studies or interventional studies in English-language, focusing on a particular subject related to oncology and published between March 2009 and March 2014 in BMJ, AoO or BjC journals were eligible for inclusion. The main outcome of the study was to be accompanied by a corresponding statistical test.

Exclusion criteria were: Phase I, II, or IV trials; descriptive studies, meta-analyses, systematic reviews, prediction model building studies, test validation studies, recommendations, letter to the editor, medico-economic studies and fundamental research studies. Studies based on censored data and / or having the survival as primary endpoints were excluded from the analysis for two reasons: first tests used (log-rank test, cox proportional hazards survival analysis) are very specific ones. Second, those studies are generated in a purpose of drug development and authorization: in this case, data produced are most carefully verified by safety administrations. Those controls are far beyond those performed in a simple publication process. So included them would bias the results. This subject should be reported on its own, elsewhere.

From identified articles, the screening consisted to realize a first articles selection, on title and on abstract. Then, the potentially eligible articles were selected on full text. Two independent readers (RR and VD) made separately a selection with those criteria. Then the two remaining selections were compared and articles with a conflicting selection results were reviewed by a third reader, a senior methodologist (AB).

Data collection

One author (RR) extracted the following data from included studies and a second author (MO) checked the extracted data: Name of publishing journal, year of publication, number of patients included, cancer studied, study type, description of outcome measures in Methods (interventional studies), definition of variables in Methods (observational studies), description of statistical test related to main endpoint/variables in Methods, type and name of statistical test, same name of statistical test used for main endpoint/variables assessment in Results, as compared to Methods, name of this test in Results, significant results for main endpoint, concordant conclusion on the main endpoint result.

Data measurement

Two variables were constructed to describe the quality of outcomes reporting: the first one is the "variables description" for the observational studies, the second one "primary outcome description" for interventional studies. For observational studies, the authors used STROBE recommendations [4] to carry out the scoring of the variable. For criterion "definition of variables in Methods", the authors noted the answer "yes" only if there was a descriptive and exhaustive definition of all the variables considered and included in the analysis, such as outcomes, exposures, predictors, potential confounders and potential effect modifiers. A descriptive and exhaustive definition of diagnostic criteria was also needed for obtaining a "yes" answer. In all other cases, the authors noted the "no" answer for this criterion. In Results section, authors recorded if a statistically significant association between disease and the exposures was reported.

Concerning interventional studies, authors relied on the CONSORT recommendations [10] to score the quality of reporting for primary endpoint. If all pre-specified outcome measures, were identified and exhaustively defined, the authors noted the answer "yes" to criterion "description of outcome measures in Methods". In all other situations, this criterion has been side "no" by the authors. It was also studied whether results for primary end point have met statistical significance.

Concerning consistency for statistical tests between Methods and Results parts in included studies, only one test per article was considered: it was the statistical test performed to assess the main endpoint of each study. First, all included studies were categorized into 3 groups in order to get a categorical variable: One group with perfect agreement for statistical test that authors considered as the most suitable methodology. A second group in which name of test was described in Methods section but not reminded in Results section, that was considered as an acceptable methodology. The third group included all articles with obvious discrepancy between the two sections, and was consequently considered as manuscripts with major methodological problem.

Secondly, a qualitative description of the causes of the discrepancies observed for the articles selected in the third group was made, in order to explore it.

Statistical methods

For the first two objectives, variables collected and constructed were described with %, Frequency and number of missing data. For the third part of the analysis, in order to identify studies' and Journals' factors associated with a complete statistical test consistency, tests have been performed on a recombined categorical variable: All studies with a perfect consistency for statistical test were classified in a "1" group and all other studies were classified in a "0" group. Variables tested in the univariable analysis were: Name of the Journal, Year of publication, Sample Size, tumor Site, Type of Study, Concordant Conclusion on the Main EndPoint, Type of Test, Main EndPoint described in methods. Odds Ratios (OR), their 95% confidence intervals and two-sided P-values (Likelihood Ratio-test: LR-test) were estimated by conditional logistic regression. A multivariable model was planned to be realized with every available covariates, since there was only 8 of them, against 826 articles to be analyzed. Analyses were carried out with the R software, version 3.1.2 (<http://www.R-project.org>). α -risk was set at 5%. Significance level was set for p-values ≤ 0.05 .

Results

Literature search

The search in BMJ, AoO, and BJC archives resulted in 5221 identified records. After title and abstract screening, 970 full text articles were assessed for eligibility. Finally, 826 studies met inclusion and exclusion criteria and were included in analysis (Fig 1).

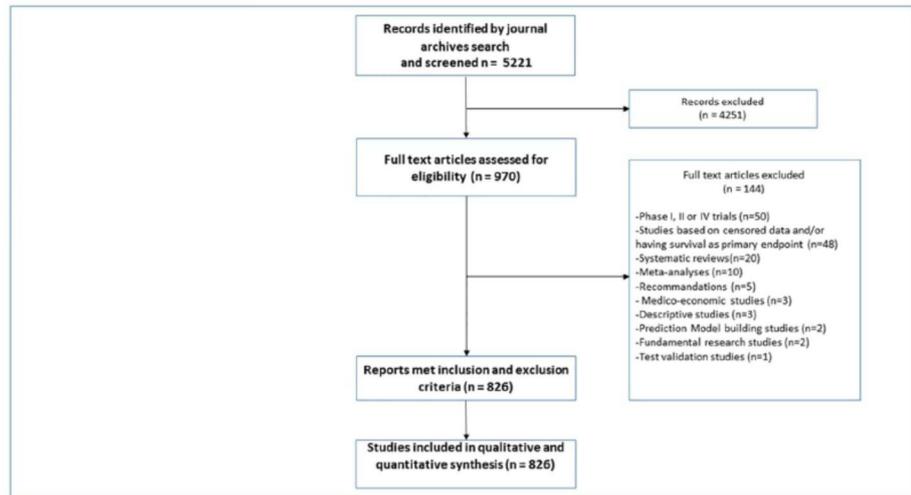


Fig 1. Flow chart: selection of studies process.

doi:10.1371/journal.pone.0164275.g001

Characteristics of selected studies

Among the 826 studies included in systematic review (S1 file), 698 (84.5%) corresponded to observational studies and 128 (15.5%) were interventional studies, comprising 29 (3.5%) phase III randomized controlled trials. 401 (48.5%) of studies were published in *Annals of Oncology*, 378 (45.8%) in *British Journal of Cancer* and 47 (5.7%) in *British Medical Journal*. When only one tumor site was studied, main reported localizations were: Breast cancer (21.8% of studies), colon and/or rectum cancer (11.3% of studies), gynecological cancer (7% of studies, Table 1).

Reporting of variables and outcomes definitions

Variables were described in Methods section for all studies. A significant association between exposure and disease was observed for 618 studies (88.6%) in observational studies. Primary endpoint was clearly detailed in Methods section for 109 interventional studies (85.2% of cases). Result for the primary endpoint was statistically significant for 87 studies (68% of cases). A concordant conclusion on the main result was observed in 97.7% of cases.

Reporting of statistical test

In observational studies, for the main analysis, test was parametric in 88.8% of cases (620 studies). Among these tests, logistic regression was the most used (416 studies, 59.6% of cases), followed by the chi-square test (70 studies, 10% of cases). Consistency of reporting for statistical test between methods and results section was evaluated as perfect for 295 studies (42.2% of cases). Name of test was not reminded in Results section for 385 studies (55.2% of cases) and an obvious discrepancy for test reporting was observed in 18 studies (2.6% of cases, Table 2).

For the interventional studies, the test chosen was mainly parametric (89 studies, 69.5% of cases). The tests most frequently used were: chi-square test (21 studies, 16.4%), logistic regression (19 studies, 14.8%), t-test (16 studies, 12.5%). A perfect consistency of reporting for

Table 1. Characteristics of Studies Included in Analysis (N = 826).

Characteristics	AoO (N = 401)	BjC (N = 378)	BMJ (N = 47)	Total, N (%)
Type of study				
Observational	334	326	38	698 (84.5)
Interventional	67	52	9	128 (15.5)
Sample size, n ^o of patients				
Median	331	960	4439	608
Interquartile range	119–1136	254–4569	1157–34352	160–2574
Year of publication				
2009	52	62	9	123 (14.9)
2010	62	61	15	138 (16.7)
2011	82	73	7	162 (19.6)
2012	112	72	7	191 (23.1)
2013	72	91	5	168 (20.3)
2014	21	19	4	44 (5.3)
Tumor Site				
One site studied	302	308	35	645 (78.1)
Breast	104	69	7	180 (21.8)
Colon/Rectum	33	54	6	93 (11.3)
Gynecological*	18	29	11	58 (7)
Leukemia/Lymphoma	35	17	0	52 (6.3)
Prostate	11	23	6	40 (4.8)
Lung	25	14	0	39 (4.7)
Stomach/Oesophagus	9	18	2	29 (3.5)
Renal	10	12	0	22 (2.7)
Pancreas	10	9	0	19 (2.3)
Head and Neck	9	9	0	18 (2.2)
Other	38	54	3	95 (11.5)
Two or more sites studied	100	69	12	181 (21.9)

AoO: Annals of Oncology; BjC: British Journal of Cancer; BMJ: British Medical Journal.

* Corresponds to the following sites: Cervix, Ovarian, Endometrium, Vagina, vulva.

doi:10.1371/journal.pone.0164275.t001

statistical test between Methods and Results sections was observed for 43 studies (33.6% of cases). Name of test was not reminded in Results section for 80 studies (62.5% of cases) and obvious discrepancy for test reporting was noted in 5 studies (3.9%, [Table 3](#)).

Discrepancy in statistical tests (qualitative assessment)

Among the 23 articles classified in the “discrepancy in statistical methods reporting” groups, no articles were published in the BMJ. 8 articles were published in BJC and 15 were published in AoO. [Table 4](#) reports the causes of discrepancy for observational studies. To illustrate the different types of discrepancies encountered in these articles, 2 types of discrepancies examples are given. First type: for some articles the tests were mentioned neither in the Methods section nor in the Results section. Second Type: tests reported in Methods section could not have given results presented in Results section (In one article method section reported solely that comparisons between groups were made using chi-square statistics. But in results Odds Ratio with a 95%CI were given).

Table 2. Reporting of variables and statistical test in observational studies (N = 698).

Reporting Item	AoO	BjC	BMJ	Total, N (%)
Variables described in Methods				
Yes	334	326	38	698 (100%)
No	0	0	0	0 (0%)
Significant association disease/exposures				
Yes	300	286	32	618 (88.6%)
No	34	39	6	79 (11.3%)
NA ¹	0	1	0	1 (0.1%)
Concordant conclusion on the main result				
Yes	329	320	37	686 (98.3%)
No	5	6	1	12 (1.7%)
Test described in Methods				
Yes	324	320	38	682 (97.7%)
No	10	6	0	16 (2.3%)
Type of test				
Parametric	284	298	38	620 (88.8%)
Nonparametric	40	22	0	62 (8.9%)
ND ¹	10	6	0	16 (2.3%)
Name of test				
Logistic regression	173	217	26	416 (59.6%)
Chi-square	44	23	3	70 (10%)
Poisson regression	11	13	6	30 (4.3%)
Wilcoxon / Kruskal-Wallis	17	12	0	29 (4.2%)
Linear regression	11	14	2	27 (3.9%)
t-test	13	11	0	24 (3.4%)
Fisher exact test	16	8	0	24 (3.4%)
Other ²	39	22	1	62 (8.9%)
ND ¹	10	6	0	16 (2.3%)
Statistical test consistency between Methods and Results sections				
Perfect Agreement	155	127	13	295 (42.2%)
Name of test not reminded in Results	169	191	25	385 (55.2%)
Obvious discrepancy	10	8	0	18 (2.6%)

AoO: Annals of Oncology; BjC: British Journal of Cancer; BMJ: British Medical Journal.

¹ ND: Not Described in the Methods section.

²Other: kappa test, ANOVA test, Pearson correlation test, Spearman test.

doi:10.1371/journal.pone.0164275.t002

Factors associated to discrepancy in statistical tests reporting

338 studies (40.9%) were included in group “1” and 488 studies (59.1%) were classed in group “0”. In univariable analysis, the three following trial characteristics were associated with perfect consistency for statistical test, with a p. value ≤ 0.2: Sample size, Type of study and Name of journal (Table 5).

In multivariable analysis, Sample size (number of included patients), remained an independent factor associated with consistency: aOR (adjusted on Name of the Journal, Year of publication, tumor Site, Type of Study, Concordant Conclusion on the Main EndPoint, Type of Test and Main EndPoint described in methods) for group 3 compared to group 1 was equal to 0.58 [0.38–0.88] (P value = 0.014, LR-test, Table 6).

Table 3. Reporting of variables and statistical test in interventional studies (N = 128).

Reporting Item	AoO (N = 67)	BjC (N = 52)	BMJ (N = 9)	Total, N (%)
Primary endpoint described in Methods				
Yes	56	44	9	109 (85.2)
No	11	8	0	19 (14.8)
Significant results for primary endpoint				
Yes	38	41	8	87 (68)
No	25	9	1	35 (27.3)
NA ¹	4	2	0	6 (4.7)
Concordant conclusion on the main result				
Yes	64	52	9	125 (97.7)
No	3	0	0	3 (2.3)
Test described in Methods				
Yes	62	49	9	120 (93.8)
No	5	3	0	8 (6.2)
Type of test				
Parametric	44	38	7	89 (69.5)
Nonparametric	18	11	2	31 (24.2)
ND ²	5	3	0	8 (6.3)
Name of test				
Chi-square	12	7	2	21 (16.4)
Logistic regression	10	8	1	19 (14.8)
t-test	9	7	0	16 (12.5)
Wilcoxon / Kruskal-Wallis	6	6	1	13 (10.2)
Fisher exact test	7	5	1	13 (10.2)
Linear regression	2	8	2	12 (9.4)
Poisson regression	1	0	2	3 (2.3)
Other ³	15	8	0	23 (18)
ND ²	5	3	0	8 (6.2)
Statistical test consistency between Methods and Results sections				
Perfect Agreement	21	19	3	43 (33.6)
Name of test not reminded in Results	42	32	6	80 (62.5)
Obvious discrepancy	4	1	0	5 (3.9)

AoO: Annals of Oncology; BjC: British Journal of Cancer; BMJ: British Medical Journal.

¹ NA: Not Assessable.

² ND: Not Described in the Methods section.

³ Other: binomial test, ANCOVA test, ANOVA test, Pearson correlation test.

doi:10.1371/journal.pone.0164275.t003

Discussion

Among all 826 articles in the present systematic review, 698 (84.5%) were observational studies and there was a good quality of reporting of variables and statistical tests in Methods parts (100% and 97.7% of cases, respectively). However, only 295 observational studies (42.2% of cases) had a perfect consistency of statistical tests between Methods and Results part. The same observations can be highlighted for interventional studies: A good reporting of the primary endpoint in 85.2% of cases and of the statistical test in the Methods section in 93.8% of cases. Similarly, a perfect consistency for statistical tests between methods and results sections was found in only 43 studies (33.6% of cases). This can be explained by a great proportion of

Table 4. Description of causes of discrepancy for statistical tests in observational studies (N = 18).

Causes	Number of articles	Journals
tests only described in results	6	2 BJC, 4 AoO
Tests described in methods and results, but a test in results is added (unexpected?)	3	3 AoO
Tests described neither in methods nor in results	4	2 BJC, 2 AoO
Tests described neither in methods nor in results, but cited in a reference (in S1 File or bibliography)	2	1 BJC, 1 AoO
Test reported in methods and results is eventually the same but reported with two different names (eg: Mann Whitney vs Wilcoxon)	1	1 AoO
Test reported in methods does not allow to get the results reported	1	1 AoO
Test described only in methods section	1	1 BJC

AoO: Annals of Oncology; BJC: British Journal of Cancer.

doi:10.1371/journal.pone.0164275.t004

studies, in which statistical tests are not clearly reported in Results section (54.9% of observational studies and 62.5% of interventional studies) and consistency cannot be assessed.

In fact, each journal establishes and publishes their specific requirements for data analysis, and there is no consensus for this aspect of peer review: Some journal editors currently request a statistical analysis of trial data by an independent biostatistician before accepting studies for publication. Others ask authors to say whether the study data are available to third parties to view and/or use/reanalyze, while still others encourage or require authors to share their data with others for review or reanalysis (*Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals, 2013*). Another element to be considered, is the poor communication, concerning statistical methods, between statistician(s) and leading author(s). Consequently, names of statistical tests utilised in Results section are inconsistently reported in publications. So, a significant proportion of studies lacked transparency for description of the statistical test in the Results section and this may be evidence of a gap in current reporting guidelines [4, 13]. STROBE and CONSORT statements recommend a detailed and complete description of the statistical test chosen for the primary endpoint in Methods section. However, nothing is said about reporting of a reminder of the statistical test used in Results section, and currently, readers of observational and interventional studies are not always able to verify the consistency for statistical test. Consequently, reproducibility of such studies cannot be assured. In literature, some articles and books have already demonstrated the importance of the reporting for statistical analysis and statistical results in high quality scientific publications [14,15]. We observed that in 18 observational studies (2.6%) tests announced were not performed. An explanation could be that although authors actually performed all analyses mentioned in the method section, they didn't report all of them in the result section because results were of no interest. Yet such information and withdrawal decision should still be mentioned. The results of this study might suggest that a systematic statistical checking should be recommended during the reviewing process, in order to raise such discrepancy between tests reported in methods and results section.

The qualitative analysis of the discrepancies highlighted the possible failure of the peer review system: in some cases, simple CONSORT and STROBE recommendations have not been followed in reporting (tests reported in statistics, rather than in methods section), in other cases, some rough mistakes have not been corrected by an attentive reviewing. In conclusion, statistical aspects of studies in oncology could be more carefully described before submission

Table 5. Analysis by univariable logistic regression of perfect statistical tests consistency for selected variables.

Variable	Univariable Analysis	
	OR (95%CI)	p value ¹
Journal		0.2
<i>Annals of Oncology</i>	1	
<i>British Journal of Cancer</i>	0.8 [0.6–1.07]	
<i>British Medical Journal</i>	0.66 [0.35–1.24]	
Year of publication		0.452
2009	1	
2010	1.18 [0.72–1.93]	
2011	0.82 [0.51–1.33]	
2012	1.03 [0.65–1.62]	
2013	0.86 [0.54–1.38]	
2014	0.64 [0.31–1.32]	
Sample size, n¹ of patients		0.005
Grp1*: 10 ≤ n < 160	1	
Grp2*: 160 ≤ n < 2.57*10 ³	0.92 [0.66–1.29]	
Grp3*: 2.57*10 ³ ≤ n < 1.09*10 ⁷	0.55 [0.37–0.82]	
Tumor Site		0.42
Breast	1	
Colon/Rectum	1.02 [0.61–1.71]	
Gynecological	1.03 [0.56–1.9]	
Leukemia/Lymphoma	1.69 [0.91–3.14]	
Prostate	1.69 [0.85–5.36]	
Lung	1.3 [0.65–2.63]	
Stomach/Oesophagus	1.81 [0.82–3.98]	
Renal	0.96 [0.38–2.42]	
Pancreas	2.32 [0.89–6.05]	
Head and Neck	2.11 [0.79–5.6]	
Other	1.08 [0.65–1.79]	
Type of study		0.132
Observational	1	
Interventional	0.75 [0.48–1.16]	
Phase III RCT ²	0.52 [0.23–1.19]	
Concordant conclusion on the MEP³ result		0.245
No	1	
Yes	1.93 [0.61–6.1]	
Type of test		0.43
Parametric	1	
Nonparametric	0.83 [0.54–1.28]	
MEP³ described in Methods		0.92
No	1	
Yes	0.95 [0.38–2.39]	

* Group 1,2 and 3.

¹ P value calculated from Likelihood Ratio-test.

² Phase III Randomised Controlled Trials.

³ MEP: Main End Point.

doi:10.1371/journal.pone.0164275.t005

Table 6. Analysis by multivariate logistic regression of perfect statistical tests consistency.

N = 628	adj. OR(95%CI)	p value [†]
Journal		
Annals of Oncology	1	
British Medical Journal	0.8 (0.36,1.78)	0.468
British Journal of Cancer	0.8 (0.56,1.15)	
Sample size, N^o of patients		
Grp1*: 10ze, N ^o o	1	0.009
Grp2*: 160≤6009: 10ze,	1.02 (0.67,1.54)	
Grp4*: 2.57*10 ³ ≤ n< 1.09*10 ⁷	0.52 (0.31,0.89)	
Year of publication		
2009	1	
2010	1.39 (0.77,2.48)	
2011	0.91 (0.52,1.6)	0.691
2012	1.1 (0.64,1.9)	
2013	0.99 (0.55,1.77)	
2014	0.84 (0.35,2.01)	
Tumor Site		
Breast	1	
Colon/Rectum	1.18 (0.68,2.03)	
Gynecological	1.27 (0.66,2.43)	0.253
Renal	1.01 (0.39,2.63)	
Leukemia/Lymphoma	1.63 (0.84,3.15)	
Lung	1.63 (0.77,3.46)	
Head and Neck	2.4 (0.87,6.58)	
Other	1.16 (0.68,1.98)	
Pancreas	3.08 (1.13,8.36)	
Prostate	2.16 (1.03,4.52)	
Stomach/Oesophagus	1.92 (0.85,4.34)	
MEP described in Methods		
No	1	
Yes	0.97 (0.28,3.39)	0.963
Type of study		
Observational	1	
Interventional	0.69 (0.38,1.23)	0.206
Phase III RCT	0.51 (0.19,1.35)	
Type of test		
Non-parametric	1	
Parametric	0.64 (0.38,1.06)	0.084
Concordant conclusion on the MEP result		
No	1	
Yes	1.4 (0.39,5.05)	0.61

* Group 1,2 and 3.

[†] P value calculated from Likelihood Ratio-test.

doi:10.1371/journal.pone.0164275.t006

for publication and adding of a statistical reviewer in medical journals, which improve quality of reporting [16], should be mandatory.

Results of the multivariable analysis have revealed that the sample size (number of patients included) is an independent factor associated with perfect statistical test consistency ($p = 0.009$). In studies with small sample sizes, the parametric assumptions are not always applicable and, therefore, the authors pay particular attention to the types of statistical tests used (parametric or nonparametric) and report them more frequently, both in Methods section and in Results section. As regards the statistical analyses section, 88.8% of observational studies and 69.5% of interventional studies used parametric tests, which gives us to understand that the majority of authors put forward a normal assumption for the distribution of their variables.

Our systematic review has several limitations. First, analysis focused on 3 journals. The results should therefore be confirmed, although selection of these three journals is fairly representative of literature in oncology. We must also remember that these results mainly concerns reporting of observational studies and we cannot draw any definite conclusion on included interventional studies (128 studies, 15.5%). A further work should be to explore whether the methodology used in those articles is appropriate, and in accordance with the actual statistical guidelines. Such work can only be undertaken with the help of a large sample of leading experts in statistics, in order to get to a consensual definition of appropriateness. Another issue raised by this review is the problem of the multiplicity in statistical tests performed in observational studies: we reported the frequency of studies where a significant association was found between exposure and disease but we did not measure in those studies the quantity of statistical multiplicity and thereby the risk of spurious conclusions it would have led to. The issue of statistical multiplicities in observational oncological studies remains to be address elsewhere.

In conclusion, in Methods section, our results show a 100% frequency of reporting for variables in observational studies and a 85% frequency for Primary Endpoint in interventional studies. A discrepancy between tests reported in methods and results sections was identified in 23 articles, whose 18 were observational studies. Current guidelines, like STROBE or CONSORT, do not yet take into account this aspect of reporting and thus we encourage authors and peer reviewers to carefully verify consistency of statistical tests.

Supporting Information

S1 File. Articles extraction form data base.

(XLS)

S2 File. PRISMA checklist.

(DOCX)

Author Contributions

Conceptualization: AB FC NM.

Data curation: RR FT MO.

Formal analysis: RR FT MO.

Investigation: VD RR MO.

Methodology: AB FC RR.

Project administration: AB.

Resources: FC NM.

Software: RR FC MO VD.

Supervision: AB FC NM.

Validation: AB.

Visualization: VD MO FT FC NM AB.

Writing – original draft: RR.

Writing – review & editing: VD MO FT FC NM AB.

References

1. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352:609–13. doi: [10.1016/S0140-6736\(98\)01085-X](https://doi.org/10.1016/S0140-6736(98)01085-X) PMID: [9746022](https://pubmed.ncbi.nlm.nih.gov/9746022/)
2. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; 276:637–9. PMID: [8773637](https://pubmed.ncbi.nlm.nih.gov/8773637/)
3. Hopewell S, Dutton S, Yu L-M, Chan A-W, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010; 340:c723. doi: [10.1136/bmj.c723](https://doi.org/10.1136/bmj.c723) PMID: [20332510](https://pubmed.ncbi.nlm.nih.gov/20332510/)
4. Vandenberghe JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007; 4:e297. doi: [10.1097/EDE.0b013e3181577511](https://doi.org/10.1097/EDE.0b013e3181577511) PMID: [17941715](https://pubmed.ncbi.nlm.nih.gov/17941715/)
5. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6:e1000097. doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097) PMID: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)
6. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Eur J Cancer* 2005; 41:1690–6. doi: [10.1038/sj.bjc.6602678](https://doi.org/10.1038/sj.bjc.6602678) PMID: [16043346](https://pubmed.ncbi.nlm.nih.gov/16043346/)
7. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (2013). [cited 15/03/2015] Available: www.icmje.org
8. Hopewell S, Collins GS, Boutron I, Yu L-M, Cook J, Shanyinde M, et al. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. *BMJ* 2014; 349:g4145. doi: [10.1136/bmj.g4145](https://doi.org/10.1136/bmj.g4145) PMID: [24986891](https://pubmed.ncbi.nlm.nih.gov/24986891/)
9. Papatheanasiou AA, Zintzaras E. Assessing the quality of reporting of observational studies in cancer. *Ann Epidemiol* 2010; 20:67–73. doi: [10.1016/j.annepidem.2009.09.007](https://doi.org/10.1016/j.annepidem.2009.09.007) PMID: [20006277](https://pubmed.ncbi.nlm.nih.gov/20006277/)
10. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg* 2012; 10:28–55. doi: [10.1016/j.ijsu.2011.10.001](https://doi.org/10.1016/j.ijsu.2011.10.001) PMID: [22036893](https://pubmed.ncbi.nlm.nih.gov/22036893/)
11. CiteFactor- Academic Scientific Journals- Journal Impact Factor List 2014 [cited 01/04/2015]. Available: <http://www.citefactor.org/journal-impact-factor-list-2014.html>
12. Inserm: Institut national de la santé et de la recherche médicale-Inserm sites- Biblioinserm- Reserved access [cited 01/04/2015]. Available: <http://biblioinserm.inist.fr/>
13. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340:c332. doi: [10.1136/bmj.c332](https://doi.org/10.1136/bmj.c332) PMID: [20332509](https://pubmed.ncbi.nlm.nih.gov/20332509/)
14. Sarter M, Fritschy J-M. Reporting statistical methods and statistical results in EJM. *Eur J Neurosci* 2008; 28:2363–4. doi: [10.1111/j.1460-9568.2008.06581.x](https://doi.org/10.1111/j.1460-9568.2008.06581.x) PMID: [19087166](https://pubmed.ncbi.nlm.nih.gov/19087166/)
15. Lang TA, Secic M. How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors and Reviewers, 2nd Edn. ACP Press 2006, Philadelphia, PA.
16. Cobo E, Selva-O'Callaghan A, Ribera J-M, Cardellach F, Dominguez R, Vilardell M. Statistical reviewers improve reporting in biomedical articles: a randomized trial. *PLoS ONE* 2007; 2:e332. doi: [10.1371/journal.pone.0000332](https://doi.org/10.1371/journal.pone.0000332) PMID: [17389922](https://pubmed.ncbi.nlm.nih.gov/17389922/)

4. Quality of reporting in oncology phase II trials : A 5-Year assesment through systematic review

4-1 Introduction

Le quatrième article original a été réalisé à la suite des conclusions de l'article 2. Il a été signalé précédemment que l'essai de phase II est une étape cruciale dans le développement de molécules innovantes en oncologie, que le taux d'attrition demeure particulièrement élevé en cancérologie lors de ce développement de nouvelles thérapeutiques et qu'une des causes pouvant expliquer ce taux d'attrition élevé serait l'utilisation récurrente dans les essais d'une méthodologie inadaptée à une étude de phase II de traitement anticancéreux. Afin de confirmer cette hypothèse, l'évaluation du reporting reste la manière la plus directe pour décrire et évaluer la méthodologie utilisée dans les essais de phase II d'oncologie. Le reporting reste un élément clé d'évaluation de la qualité d'un essai, car il représente ce qui est directement disponible pour le lecteur.

Il a été signalé dans la revue de littérature précédente que l'évaluation de la qualité du reporting des essais de phase II en oncologie avait été très peu étudiée (seulement 13 études). Un score d'évaluation de la qualité globale du reporting n'avait été utilisé que dans 2 études. Enfin, la majorité des articles qui avaient été inclus dans la revue de littérature précédente avaient analysé des essais de phase II publiés avant 2010, c'est-à-dire avant l'avènement de nouvelles classes thérapeutiques en cancérologie comme les thérapies moléculaires ciblées ou l'immunothérapie.

L'objectif principal de ce quatrième article original a donc consisté à réaliser une revue systématique de la littérature, décrivant et analysant la qualité du reporting des articles d'essais de phase II d'oncologie et publiés dans trois journaux médicaux depuis 2010. L'évaluation de la qualité du reporting a été réalisée à l'aide d'un score intitulé Key Methodological Score (ou score KMS). L'objectif secondaire a été d'identifier d'éventuel(s) facteur(s), présent(s) dans les publications, associé(s) à une valeur maximale du score KMS.

4-2 Méthodes

Trois journaux publiant des essais d'oncologie ont été sélectionnés par tirage au sort parmi une sélection de journaux d'oncologie ayant un impact factor supérieur à 4. Les journaux finalement analysés étaient le Journal of Clinical Oncology (JCO), Annals of Oncology (AoO) et le British Journal of Cancer (BJC). Tous les articles d'essais d'oncologie de phase II et de phase II/III publiés dans ces trois journaux entre mars 2010 et février 2015, étaient incluables dans la revue systématique de littérature. Les articles suivants ont été exclus de l'analyse : Les essais de phase I/II, les réanalyses de sous-groupes de patients inclus dans des essais de phase II ou de phase II/III, les analyses groupées de plusieurs essais de phase II ou de phase II/III, les articles publiant uniquement les résultats de la partie phase III des essais de phase II/III. Les articles d'essais de phase II ou de phase II/III avec un design Bayésien ont été également exclus de cette sélection, à cause de leur méthodologie assez différente de celle utilisée dans les essais de phase II/III avec un design non Bayésien. L'analyse de la qualité de

reporting d'articles d'essais ayant choisi le design Bayésien ne peut pas être réalisée avec les mêmes scores que ceux utilisés pour les articles d'essais à design non Bayésien. L'inclusion de tels articles aurait ainsi provoqué une certaine hétérogénéité dans l'analyse.

La sélection des études a été réalisée de façon indépendante par deux auteurs (Julien Langrand Escure et Mathieu Oriol) et ces sélections ont ensuite été comparées. Chaque article discordant entre les deux méthodes de sélection a été lu par un troisième auteur (Aurélié Bourmaud), qui a pris ensuite la décision d'inclure ou non l'article dans l'analyse.

Un fichier d'extraction a été développé par une équipe pluridisciplinaire composée d'un méthodologiste (Aurélié Bourmaud), d'un statisticien (Fabien Tinquaut) et de cliniciens (Julien Langrand Escure et Romain Rivoirard). Ce fichier a été tout d'abord testé sur dix articles sélectionnés au hasard, et modifié en conséquence. Un auteur (Julien Langrand Escure) a extrait les données des articles. Les variables suivantes ont été collectées : Année de publication, site tumoral étudié, type de traitement évalué, association médicamenteuse ou non, type de thérapie en cas de traitement systémique administré, analyse de biomarqueurs (oui / non), design statistique utilisé, reporting du design statistique (oui / non), essai comparatif ou non s'il y a plus d'un bras de traitement dans l'étude, type de plan statistique, nombre d'étapes dans le plan statistique, ratio de randomisation en cas d'essai randomisé (équilibré/non équilibré), valeurs de l'erreur de type I et de type II, reporting de l'hypothèse a priori (oui / non), seuil à atteindre pour conclure dans les essais non comparatifs (rapporté/non rapporté), pourcentage de perdus de vue prévu, type de financement, analyse intermédiaire prévue (oui / non), arrêt anticipé prévu suite à l'analyse intermédiaire (oui/non), nombre de centres participant à l'étude, résultats de l'étude, type de critère de jugement principal, nombre de patients prévus a priori, inclus et analysés, le type d'analyse prévu et finalement réalisé.

Afin d'établir le score KMS pour chaque article d'étude, les trois items de reporting suivants ont été systématiquement notés pour chaque publication : Le reporting du critère de jugement principal, le reporting des éléments nécessaires justifiant le calcul du nombre de patients à inclure dans l'étude, reporting de la définition de la population évaluable. Le critère de jugement principal a été considéré comme bien rapporté s'il était clairement décrit dans la section Introduction ou dans la section Méthodes de l'article. L'item justification du calcul du nombre de patients à inclure nécessitait le reporting d'au moins l'hypothèse a priori, ainsi que les valeurs d'erreur de type I et de type II. La définition de la population évaluable était examinée et définie comme une sélection et une évaluation similaire de la population, prévue dans la section Méthodes et réalisée dans la section Résultats. Un point a été attribué pour chaque item s'il était bien rapporté dans la publication. Les scores minimaux et maximaux du KMS pouvaient être respectivement égaux à 0 et 3.

Une analyse descriptive a été réalisée : les variables qualitatives ont été décrites à l'aide de fréquences et de pourcentages. Le score KMS été décrit à l'aide de pourcentages (score = 0, 1, 2 ou 3) et de sa valeur médiane (plage interquartile ainsi que valeurs minimale et maximale). Dans un second temps, des analyses univariées et multivariées ont été réalisées afin d'identifier des variables présentes dans les articles, possiblement associées à un score

KMS maximal (KMS=3), versus tous les autres scores du KMS (KMS<3). Les variables suivantes ont été incluses dans l'analyse univariée : Journal ayant publié l'essai, année de publication, site tumoral, design statistique utilisé, reporting du design statistique, essai de combinaison de médicaments, analyse non comparative versus comparative, ratio de randomisation en cas d'essai randomisé (équilibré/non équilibré), analyse des biomarqueurs, reporting de la valeur seuil à atteindre pour conclure, type de financement, arrêt prématuré de l'essai sur une analyse intermédiaire, nombre de centres participants à l'essai, résultats de l'essai, type de critère de jugement principal, nombre de patients à inclure à priori dans l'essai, comparaison entre le nombre de patients anticipés /inclus et anticipés /analysés, type d'analyse prévue et finalement réalisée. Une régression logistique a été effectuée afin d'obtenir un odds ratio pour chaque variable, avec un intervalle de confiance de 95%. (IC_{95%}). Les variables identifiées avec une valeur de petit p inférieure ou égale à 0,2 à la suite de l'analyse univarié et ne montrant aucune corrélation, ont été incluses dans un modèle multivarié et analysées avec une régression logistique. Les variables avec une valeur de petit p inférieure à 0,05 après analyse multivariée ont été considérées comme associées de manière significative à la valeur du score KMS. Toutes les analyses statistiques ont été effectuées avec le logiciel R 3.1.1.

4-3 Résultats

Les auteurs ont identifié 574 et 606 articles d'essais de phase II ou de phase II / III potentiellement incluables, respectivement par sélection manuelle dans les archives des trois journaux sélectionnés et par recherche internet via MEDLINE avec filtrage manuel. Les 2 processus de sélection ont identifié 538 publications communes. L'analyse des articles discordants a été réalisée par un troisième auteur (Aurélien Bourmaud) et a conduit à la conservation de 30 publications supplémentaires (568 publications). Après exclusion des essais avec un design bayésien, 557 publications d'essais de phase II ou de phase II / III ont été retenues pour analyse, avec 247 articles (44,4%) publiés dans le Journal of Clinical Oncology, 218 articles (39,1%) publiés dans Annals of Oncology et 92 articles (16,5%) publiés dans le British Journal de cancer.

Les essais sélectionnés étaient essentiellement des études à un seul bras (n = 315, 56,6% des cas). 193 essais (34,6% des cas) étaient randomisés et 49 (8,8% des cas) étaient des essais à groupes multiples non randomisés. Le design de l'étude était clairement rapporté pour 395 essais (70,9% des cas), alors qu'il a été déduit mais pas clairement annoncé dans l'article pour 162 publications (29,1% des cas). La plupart des essais analysés mettaient à l'évaluation l'effet d'une chimiothérapie (78,3% des cas). Les trois sites tumoraux les plus étudiés étaient : le tractus gastro-intestinal (22,3% des cas), les maladies onco-hématologiques (15,3% des cas) et les cancers broncho-pulmonaires (12,6% des cas). 443 essais (79,5% des cas) étaient multicentriques, alors que 43 (7,7% des cas) étaient des essais monocentriques et que pour 71 essais (12,7% des cas), ces données n'étaient pas rapportées dans l'article.

La valeur médiane du score KMS était égale à 2 (interquartile Q1-Q3 : 1 - 3). Le score KMS était égal à 0, 1, 2, 3 respectivement pour 22 (3,9%), 119 (21,4%), 270 (48,5%) et 146 (26,2%) articles. Concernant l'analyse univariée, le site tumoral, le type de design statistique,

le reporting clair du design statistique, le reporting du seuil à atteindre pour conclure dans les essais non comparatifs, le nombre de centres participant à l'étude, les résultats d'essais, le nombre prévu de patients à inclure, la comparaison entre le nombre de patients anticipés / inclus et anticipés / analysés, ainsi que le type d'analyse planifiée et réalisée, étaient significativement associés avec la valeur KMS la plus élevée.

En analyse multivariée, 3 variables ont été significativement associées à la valeur du score KMS (égal à 3 contre <3): Le reporting clair versus déduit du design statistique (OR = 2.22; IC 95% [1.36-3.65]; valeur p <0.001), l'essai monocentrique versus multicentrique (OR = 0.25 [0.09 – 0.74]; valeur p = 0.012) et l'analyse menée per protocole versus selon intention de traiter (OR = 0.48; IC 95% [0.32 -0.72]; valeur p <0.001).

4-4 Discussion

La qualité du reporting pour les publications d'essais de phase II et II / III en oncologie peut être améliorée, au vu du faible nombre d'articles (26,2%des cas) qui ont correctement rapporté les trois items du score KMS, suite à cette analyse. Le reporting du design statistique, le caractère multicentrique de l'essai et l'analyse en intention de traiter sont associées de manière significative à la valeur maximale du score KMS (KMS=3/3). Par contre, aucune association significative n'a été trouvée entre le score KMS et le nom du journal de publication, l'année de publication et le type de design statistique.

Une analyse plus détaillée de cette revue systématique met en lumière plusieurs constatations : Premièrement, les trois items du score KMS ne sont pas rapportés dans les mêmes proportions : le critère de jugement principal est presque systématiquement rapporté, alors que le reporting du calcul justifiant la taille de l'échantillon est manquant dans un tiers des articles. Le reporting de la définition de la population évaluable est l'élément le moins souvent retrouvé dans les publications. La définition de la population évaluable étant l'une des 3 composantes principales du score KMS, ce taux très bas de reporting (présent dans 39.9% des publications) peut soulever des questions quant à la validité de telles études.

Le deuxième commentaire réside dans les autres éléments recueillis dans la fiche d'extraction, comme caractéristiques descriptives : Plus de la moitié des études sont monobras, un tiers des articles d'études n'ont pas rapporté clairement le design de l'étude, alors que ceci est considéré comme un élément méthodologique majeur. 20% des articles d'études de phase II et II/III ne décrivent pas d'hypothèse de recherche, 11% ne précisent pas leurs sources de financement et 43% ne définissent pas la valeur seuil à atteindre dans l'étude pour conclure Tous ces éléments de reporting sont considérés comme obligatoires dans les guidelines de reporting utilisés pour les essais de phase III (déclaration CONSORT). L'ampleur du défaut de reporting de certains items soulève également des questions quant à la validité de ces études.

À notre connaissance, seules 2 publications ont évalué le reporting des publications d'essais de phase II en oncologie à l'aide d'un score global [193,194]. Ottaiano et al. [194] ont publié un score composé de 11 items, appelé quality index. 141 essais publiés dans 5 revues d'oncologie différentes ont été analysés, sur une période de 5 ans, entre 1998 et 2002.

L'impact factor de la revue était significativement associé à un score quality index supérieur à 50/100 (valeur $p = 0,001$). Contrairement à Ottaiano et al. [194], notre analyse portant sur 3 journaux d'oncologie n'a pas pu démontrer une association entre la renommée du journal et la qualité du reporting des essais de phase II d'oncologie.

Grellety et al [193] ont publié deux scores globaux pour l'évaluation de la qualité du reporting d'articles d'essais de phase II en oncologie : l'Overall Quality Score (OQS) et le score KMS. Le premier peut être considéré, tout comme la déclaration CONSORT, comme une aide à la rédaction d'essais, pour les auteurs considérant les 44 items inclus dans ce score. Le score KMS, qui comprend seulement 3 items de reporting, est de notre point de vue, le seul outil adapté pour l'évaluation systématique du reporting dans les essais de phase II en oncologie. Les auteurs ont utilisé ces scores pour l'analyse des essais de médicaments anticancéreux, de radiothérapie et de chirurgie. Ils ont analysé 156 essais publiés dans 8 journaux d'oncologie, avec un impact factor sur 5 ans supérieur à 4. La période de sélection a été l'année 2011. Les taux de reporting de chaque élément du score KMS étaient de 68,6%, 77,6% et 33,3% respectivement pour le reporting du critère de jugement principal, le reporting de la justification du calcul du nombre de sujets à inclure et le reporting de la définition d'une population évaluable. Ces taux semblent être comparables avec nos conclusions pour les deux derniers items cités ci-dessus : 66,6% contre 77,6% pour la justification du calcul du nombre de sujets à inclure et 39,9% contre 33,3% pour la définition d'une population évaluable. Concernant le reporting du critère de jugement principal, nous avons trouvé un taux supérieur à celui de Grellety et al (90,5% par rapport à 68,6%). Une analyse statistique par régression logistique a également été réalisée par Grellety et al. Aucune variable n'était statistiquement associée à la valeur du score KMS, aussi bien pour l'analyse univariée que multivariée. Cependant, les variables explorées par les auteurs sur l'analyse univariée et multivariée n'étaient pas aussi nombreuses que dans notre revue systématique de littérature. Un impact factor supérieur à 10 et un enregistrement de l'essai sur clinicaltrials.gov étaient significativement associés à une meilleure valeur du score OQS en analyse multivariée.

La qualité du reporting a également été analysée sur un unique critère. Le reporting du design statistique semble être le critère le plus fréquemment étudié. Comme dans notre étude, Perrone et al [198] avaient montré un impact significatif du nombre de centres participants à l'essai de phase II, sur la qualité finale du reporting de l'essai. Au total, 145 publications d'études de phase II portant sur le cancer du sein, et publiées entre 1995 et 1999, avaient été analysées. Les essais multicentriques étaient significativement associés à un taux plus élevé de reporting du design statistique (OR = 3,24 ; IC_{95%} [1,47-7.15]).

Parmi les limites de l'étude, l'inclusion dans l'analyse des essais de phase II / III peut être discutable. En effet, dans l'article de Grellety et al. [193] les études de phase II / III avaient été exclues de l'analyse. Cependant, nous n'avons analysé que la partie phase II des essais de phase II / III et nous assumons le fait qu'il n'y ait aucune différence en termes d'exigence pour l'évaluation de la qualité du reporting entre la phase II seule et la partie phase II d'un essai de phase II / III. De plus, le choix du score KMS, utilisé comme score de référence pour évaluer la qualité du reporting des essais de phase II en oncologie, est

discutable. En effet, avec seulement 3 items, la limite entre un reporting d'essai considéré comme « bon » et « mauvais » est difficile à déterminer. Le choix d'utiliser ce score pour notre analyse était justifié du fait qu'il était, à notre connaissance, le seul score global publié pour l'analyse systématique de la qualité du reporting pour les essais cliniques de phase II d'oncologie. Cependant, l'amélioration et l'affinement du score KMS seraient des actions intéressantes à mettre en œuvre, afin de se rapprocher, dans sa structure, des autres scores d'évaluation du reporting.

Cette étude soutient l'hypothèse d'un reporting non optimal des essais cliniques de phase II dans les articles scientifiques publiés. Cette revue de littérature a également soulevé la question de la qualité et de la pertinence des choix méthodologiques dans la conduite de ces essais. Notre analyse enrichit le faible contenu de la littérature portant sur le sujet de la qualité du reporting des essais de phase II en oncologie. Cette étude est, à notre connaissance, la plus grande étude sur ce sujet en termes de quantité d'essais analysés, et le premier à démontrer une association statistiquement significative entre certaines caractéristiques de l'étude et la valeur du score KMS.

Etant donné qu'aucun autre score global d'évaluation du reporting n'est disponible aujourd'hui, le score KMS doit être considéré comme une référence pour l'analyse systématique de la qualité du reporting des essais de phase II en oncologie, à la fois par les auteurs et les lecteurs

L'article original en texte intégral est disponible ci-dessous (en anglais).

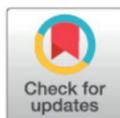
RESEARCH ARTICLE

Quality of reporting in oncology phase II trials: A 5-year assessment through systematic review

Julien Langrand-Escure^{1,2}, Romain Rivoirard³, Mathieu Oriol^{1,4}, Fabien Tinquaut^{1,4}, Chloé Rancoule², Frank Chauvin^{1,4,5}, Nicolas Magné², Aurélie Bourmaud^{1,4,5*}

1 Centre Hygiène, Public Health Department, Lucien Neuwirth Cancer Institut, Saint Priest en Jarez, France, **2** Radiotherapy Department, Lucien Neuwirth Cancer Institut, Saint Priest en Jarez, France, **3** Oncology Department, Lucien Neuwirth Cancer Institut, Saint Priest en Jarez, France, **4** INSERM 1408 CIC-EC, Saint Etienne, France, **5** EA Health Services Performance Research HESPER 7425, Lyon 1 University, Lyon, France

* Aurelie.bourmaud@icloire.fr



Abstract

Background

Phase II clinical trials are a cornerstone of the development in experimental treatments. They work as a "filter" for phase III trials confirmation. Surprisingly the attrition ratio in Phase III trials in oncology is significantly higher than in any other medical specialty. This suggests phase II trials in oncology fail to achieve their goal.

Objective The present study aims at estimating the quality of reporting in published oncology phase II clinical trials.

Data sources

A literature review was conducted among all phase II and phase II/III clinical trials published during a 5-year period (2010–2015).

Study eligibility criteria

All articles electronically published by three randomly-selected oncology journals with Impact-Factors > 4 were included: Journal of Clinical Oncology, Annals of Oncology and British Journal of Cancer.

Intervention

Quality of reporting was assessed using the Key Methodological Score.

Results

557 articles were included. 315 trials were single-arm studies (56.6%), 193 (34.6%) were randomized and 49 (8.8%) were non-randomized multiple-arm studies. The Methodological Score was equal to 0 (lowest level), 1, 2, 3 (highest level) respectively for 22 (3.9%), 119 (21.4%), 270 (48.5%) and 146 (26.2%) articles. The primary end point is almost

OPEN ACCESS

Citation: Langrand-Escure J, Rivoirard R, Oriol M, Tinquaut F, Rancoule C, Chauvin F, et al. (2017) Quality of reporting in oncology phase II trials: A 5-year assessment through systematic review. PLoS ONE 12(12): e0185536. <https://doi.org/10.1371/journal.pone.0185536>

Editor: Carlos M. Isales, Georgia Regents University, UNITED STATES

Received: November 22, 2016

Accepted: September 14, 2017

Published: December 7, 2017

Copyright: © 2017 Langrand-Escure 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 data underlying the findings are fully available within the paper and the supporting information files.

Funding: No funds.

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

systematically reported (90.5%), while sample size calculation is missing in 66% of the articles. 3 variables were independently associated with reporting of a high standard: presence of statistical design (p-value <0.001), multicenter trial (p-value = 0.012), per-protocol analysis (p-value <0.001).

Limitations

Screening was mainly performed by a sole author. The Key Methodological Score was based on only 3 items, making grey zones difficult to translate.

Conclusions & implications of key findings

This literature review highlights the existence of gaps concerning the quality of reporting. It therefore raised the question of the suitability of the methodology as well as the quality of these trials, reporting being incomplete in the corresponding articles.

Introduction

Phase II clinical trials are pivotal steps in the development of new drugs and/or new therapeutic strategies. In oncology particularly, the aim of phase II is to assess the efficacy and safety of experimental treatments on a small sample of highly selected cancer patients, in order to determine an effective and safe enough dose, which could then be further administered to a broader population. Thus, phase II trials could be considered as a filter for experimental treatments, before phase III, defined as the largest scaled step and which is an obligatory stage before treatment registration.

Clinical research in oncology has evolved significantly in the last 20 years. While oncology was a discipline with a small number of experimental treatments in development, in the early 2000's, the discovery of new systemic drugs, especially targeted therapies, and the technological progresses in radiation therapy have led to an explosion in the number of experimental treatments assessed in trials [1]. Lots of new drugs have been developed in oncology due to the urgent need to offer alternative and more efficient treatments to patients, but only a few of them have passed the stringent testing protocols of clinical research, especially during the phase II stage. Oncology is still the medical specialty with the highest attrition rate during the development of therapeutic treatment, all stages of development included. Particularly during phase III stage, oncology presents a mean attrition rate that is estimated to be 11.3% to 14% higher than the other specialties' rates [2,3], leading to the hypothesis that oncology phase II trials do not successfully fulfill their filter role. Furthermore, since financial constraints are a fundamental issue in clinical research, there is a high rationale to try improve selection capacity of oncology phase II trials because of the important global cost linked to large phase III trials.

Several hypotheses may be raised to explain this peculiar phenomenon. One being that adequate methodology, specific to drug cancer research, may not be followed by investigators. This assumption may be supported by the fact that, at the methodological level, some discussion and general recommendations, essentially concerning the choice of the statistical design, have been published, [4–16], but up until today no consensus exists. The methodology used in oncology phase II trials responds to a non-systematic scheme. One way to initiate the investigation of this hypothesis is to explore the reporting of cancer phase II trials. Reporting is a key element to assessing the quality of trials, because it represents what is available for the readers.

Concerning reporting, little data exist, and it is essentially made up of isolated methodological criteria and mainly reports the trial's statistical design [17–23]. Until 2014, no global recommendation had been published for phase II trials, such as the CONSORT guidelines for randomized controlled trials [24–26]. Since January 2014, two global scores have been validated and published [27]. They are called the Overall Quality Score and the Key Methodological Score and are used to assess the quality of reporting in oncology phase II trials [27]. Investigating reporting with a validated tool would be the first step to a broader assessment of the methodological quality in phase II trials.

The aim of the present study was to describe and assess the reporting quality of oncology phase II trials published during a 5-year period in 3 journals in oncology, following the Key Methodological Score. The secondary endpoint was to identify factors associated with a high value of the reporting score.

Methods

Study selection

Three oncology journals were randomly selected among a selection of oncology journals with a 5-year Impact Factor above 4. Journals retained were *Journal of Clinical Oncology*, *Annals of Oncology* and *British Journal of Cancer*. All phase II and phase II/III trials published during a 5-year period, between March 2010 and February 2015, in these 3 journals, were eligible. Phase I/II trials, reanalysis of the subgroup of patients included in phase II or II/III trials, pooled analysis of many phase II or II/III trials, publications reporting only the phase III part of phase II/III trials, were excluded. Phase II or II/III trials with the Bayesian design were also excluded. The choice of excluding phase II and II/III trials with the Bayesian design is justified by the difference in the methodology of these trials, compared to the non-Bayesian phase II or II/III trials. We assume that analysis of reporting in these types of trials cannot be realized with the same score as the non-Bayesian trials and that inclusion of these trials would have led to heterogeneity in the set of analyses.

The selections of studies were independently realized by two of the authors and then compared. One of the authors manually selected each phase II and phase II/III trial published during the concerned period in the journal's computerized archives, with the following key-words: "phase II" or "phase 2" or "phase II/III" or "phase 2/3" present in the title and/or in the abstract and/or in the last paragraph of the Introduction and/or in the Methods section of the article. The other author selected studies with the following electronic request in the MEDLINE database: (((("Br J Cancer."[Journal] OR "J Clin Oncol."[Journal]) OR "Ann Oncol."[Journal]) AND ("2010/03"[PDAT]: "2015/02"[PDAT])) AND Clinical Trial, Phase II[ptyp]). All the publications identified with the electronic request were then manually selected by this author in order to keep only phase II and phase II/III trials that were published in the 3 defined journals. Selected trials by both authors were conserved for analysis. Discordant articles between the two selection methods were read by a third author in order to make the final decision as to whether to include the publication for analysis. The present systematic review was not registered in PROSPERO.

Study analysis

A data extraction sheet was developed by a multidisciplinary team composed of a methodologist, a statistician and clinicians. It was then pilot-tested on ten randomly-selected included studies, and it was refined accordingly. One author extracted the data from included studies. The following variables were collected: year of publication, disease site, type of treatment, drug combination or not, type of chemotherapy in case of a chemotherapy trial, biomarker analysis

(yes/no), statistical design used, reporting of the statistical design (yes/no), comparative trial or not if more than one arm, type of statistical plan, number of stages, random ratio when randomized trial, type I and type II error values, a priori hypothesis (yes/no), cut-off point to conclude for non-comparative trials (present/missing), percentage of loss to follow-up planned, type of funding, planned intermediate analysis (yes/no), early stopping on planned intermediate analysis or not, number of centers, trial results, type of primary endpoint, number of patients anticipated, included and analyzed, type of analysis planned and realized. Trials for which the author had difficulty in completing the data collection base, were read by a second author to make a decision.

The Key Methodological Score items collection

The use of the Key Methodological Score to analyze the reporting quality of oncology phase II trials have previously been reported [27]. The three following items were systematically recorded for each publication: reporting of the primary endpoint, reporting of the elements needed to justify the sample size calculation, reporting of the definition of an evaluable population (Table 1). The primary endpoint was considered to have been reported when described in the Introduction or in the Methods section of the article. When primary endpoint was reported only in the abstract of the publication, it was considered not to have been reported. Justification of sample size calculation needed the reporting of at least the a priori hypothesis, as well as the type I and type II error values. The definition of the evaluable population was screened and defined as a similar population selection and assessment, planned in the Methods section and realized in the trial analysis. One point was attributed for each item if reported in the publication. Minimal and maximal scores could be respectively 0 and 3.

Statistical analysis

Descriptive analysis was performed: qualitative variables were described using frequencies and percentages. The Key Methodological Score was described with rate (score of 0, 1, 2 or 3) and median score (interquartile range and minimal and maximal values). Secondary, univariate and multivariate analyses were performed in order to explore variables potentially associated with the highest score of the Key Methodological Score (equal to 3) versus all other scores (lower than 3). The following variables were included in the univariate analysis: journal of publication, year of publication, site of disease, statistical design used, reporting of the statistical design, drug combination, non-comparative versus comparative analysis, type of random ratio, biomarker analysis, reporting of the cut-off point to conclude, type of funding, early stopping on intermediate analysis, number of centers, results of trial, type of primary endpoint, anticipated number of patients to include, comparison between number of patients anticipated/included and anticipated/analyzed, type of analysis planned and realized. A logistic regression was conducted in order to obtain an odds ratio for each variable, with a 95% confidence interval (CI). Variables identified with a p-value lower or equal than 0.2 in univariate analysis and showing no correlation, were included in a multivariate model and analyzed with

Table 1. Key Methodological Score as published by Grellety et al. [27].

Items	Definition
Primary end point	Clear referencing and definition of the first end point
Sample size	Hypothesis and justification of sample size clearly explained
Evaluable population	Definition of rules to consider patients as being evaluable for first and secondary end points

<https://doi.org/10.1371/journal.pone.0185536.t001>

a logistic regression. Variables with a p-value lower than 0.05 after multivariate analysis were considered as significantly associated with the the Key Methodological Score value. All the statistical analyses were conducted with R 3.1.1 software.

Results

Study selection

Authors identified 574 and 606 phase II or II/III trials of interest, respectively through manual selection in the journal's computerized archives and by electronic request in the MEDLINE database plus manual screening. The 2 selection processes identified 538 publications in common. The analysis of discordant articles realized by a third author lead to the conservation of 30 more publications (568 publications). After exclusion of the Bayesian design trials, 557 phase II or II/III trials were retained for analysis with respectively 247 (44.4%), 218 (39.1%) and 92 (16.5%) articles in *Journal of Clinical Oncology*, *Annals of Oncology* and *British Journal of Cancer*. The whole study process selection is described in the flow-chart (Fig 1).

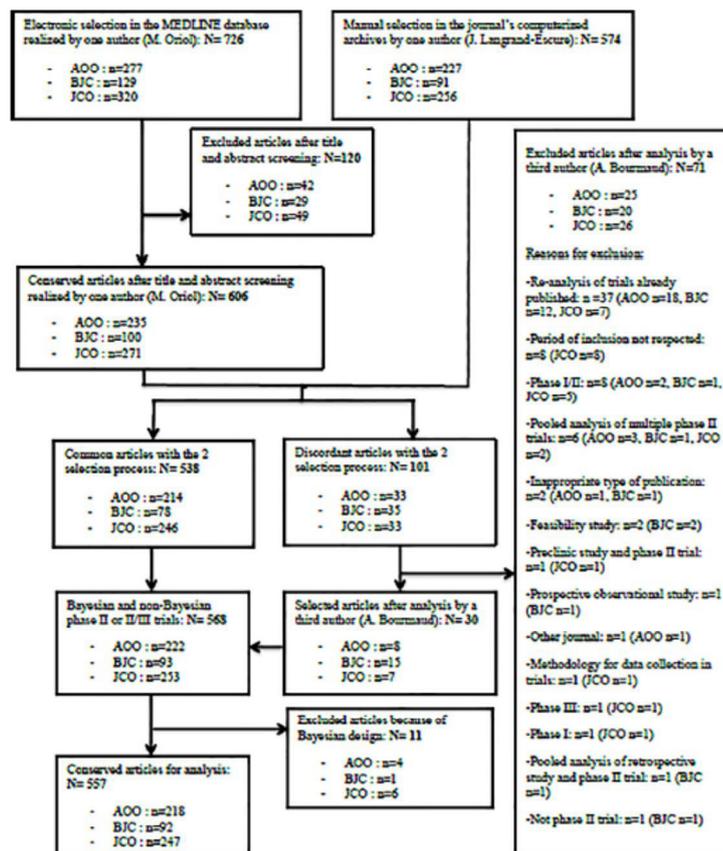


Fig 1. Flowchart.—Flowchart summarizing the selection process of the phase II trials for analysis (AOO: *Annals of Oncology*; BJC: *British Journal of Cancer*; JCO: *Journal of Clinical Oncology*).

<https://doi.org/10.1371/journal.pone.0185536.g001>

Study characteristics

Selected trials were essentially single-arm studies ($n = 315$, 56.6%). 193 (34.6%) trials were randomized and 49 (8.8%) were multiple-arm, non randomized trials. Study design was clearly reported for 395 trials (70.9%), whereas it was deduced but not clearly announced for 162 publications (29.1%). Most of the trials that were analyzed were testing chemotherapy (78.3%).

The three sites of disease with which we were most concerned, were gastrointestinal tract (22.3%), hematological disease (15.3%) and lung cancer (12.6%). 443 (79.5%) trials were multi-center trials, whereas 43 (7.7%) were single-center trials and for 71 (12.7%) trials this data was not reported. All study characteristics are detailed in [Table 2](#).

Key Methodological score

The median value of the Key Methodological Score was 2 (interquartile range: 1–3). Key Methodological Score was equal to 0, 1, 2, 3 respectively for 22 (3.9%), 119 (21.4%), 270 (48.5%) and 146 (26.2%) trials. Detailed scores for each item of the Key Methodological Score are described in [Table 3](#). Concerning univariate analysis, disease site, type of statistical design, reporting of statistical design, cut-off point to conclude, number of center, trial results, anticipated number of patients to include, comparison between number of patients anticipated/included and anticipated/analyzed, as well as type of analysis planned and realized, were significantly associated with the highest KMS value. Concerning multivariate analysis, 3 variables were significantly associated with the Key Methodological Score value (equal to 3 versus <3): reporting of statistical design (OR = 2.22; IC 95% [1.36–3.65]; p -value <0.001), single center trial versus multi-center trial (OR = 0.25 [0.09–0.74]; p -value = 0.012) and analysis conducted in intention to treat versus per-protocol (OR = 0.48; IC 95% [0.32–0.72]; p -value <0.001). The results of the multivariate analysis are summarized in [Table 4](#).

Discussion

The quality of reporting in oncology phase II and II/III trials can be improved according to the low number of clinical trials (26.2%) that reported all three items of the Key Methodological Score in our set of analysis. The reporting of statistical design, multicenter trial and trial analysis conducted in intent-to-treat are significantly associated with the highest value of the Key Methodological Score (equal to 3) in this analysis of 557 oncology phase II and II/III trials. On the other hand, no significant association was found between the Key Methodological Score and journal of publication, year of publication and type of statistical design. A further analysis of this review brought to light several issues. First, the three items of the Key Methodological Score are not reported in the same proportions: the primary end point is almost systematically reported, while sample size calculation is missing in one third of the articles. The definition of population is the least reported item, with more studies failing to report such information, than studies reporting it. The definition of the population being one of the 3 core components of the Key Methodological Score, this very low rate may raise questions concerning the validity of such studies. The second issue results from the other items collected in the extraction sheet as descriptive characteristics: more than half of the studies are single-armed, one third of the studies have failed to report the study design which could be considered as a major methodological requirement, 20% are not cited as being multicenter, 20% do not provide a research hypothesis, 11% do not provide their funding sources and 43% do not report their cut off point. All those reporting items are considered as mandatory in other study design reporting guidelines. The extent of the defect in such items also raises questions concerning the validity of these studies.

Table 2. Characteristics of phase II and II/III trials published in the 3 journals selected (N = 557).

Characteristics	Number of articles (%)		
Journal title	Journal of Clinical Oncology	247 (44.4)	
	Annals of Oncology	218 (39.1)	
	British Journal of Cancer	92 (16.5)	
Year of publication	2010	107 (19.2)	
	2011	120 (21.5)	
	2012	107 (19.2)	
	2013	125 (22.4)	
	2014	81 (14.5)	
	2015	17 (3.1)	
Type of treatment	Intravenous chemotherapy	319 (57.3)	
	Non intravenous chemotherapy	117 (21)	
	Radiochemotherapy	51 (9.2)	
	Hormonal therapy	11 (2)	
	Immunotherapy	11 (2)	
	Radiation therapy	6 (1.1)	
	Other	42 (7.5)	
Drug combination (*)	Yes	312 (62.4)	
	No	188 (37.6)	
Type of chemotherapy (*)	Combination	193 (39.5)	
	Cytotoxic	148 (30.3)	
	Targeted therapy	103 (21.1)	
	Monoclonal antibody	36 (7.4)	
	Non assessable	9 (1.8)	
Type of neoplasm	Gastrointestinal tract	124 (22.3)	
	Hematology	85 (15.3)	
	Lung	70 (12.6)	
	Breast	69 (12.4)	
	Urology	51 (9.2)	
	Gynecology	38 (6.8)	
	Head and neck	32 (5.7)	
	Skin	23 (4.1)	
	Sarcoma	17 (3.1)	
	Many sites	12 (2.2)	
	Brain	11 (2)	
	Other	25 (4.5)	
	Type of study	Phase II study	554 (99.5)
		Phase II/III study	3 (0.3)
Study design	Single arm	315 (56.6)	
	Randomized trial	193 (34.6)	
	Multiple arm non randomized trial	49 (8.8)	
Reporting of the study design	Yes	395 (70.9)	
	No	162 (29.1)	
Randomized trial (*)	Comparative	117 (60.6)	
	Non-comparative	76 (39.4)	
Multiple arm non randomized trial (*)	Comparative	6 (12.2)	
	Non-comparative	43 (87.8)	

(Continued)

Table 2. (Continued)

	Characteristics	Number of articles (%)
Random ratio (*)	Well-balanced	140 (72.5)
	Non-well-balanced	19 (9.8)
	Non assessable	34 (17.6)
Type I error value	<5%	28 (5)
	5%	210 (37.7)
]5–10%[132 (23.7)
	>10%	26 (4.7)
	Non assessable	161 (28.9)
Power value	<80%	15 (2.7)
	80%	157 (28.2)
]80–90%[50 (9)
	90%	141 (25.3)
	>90%	33 (5.9)
	Non assessable	161 (28.9)
A priori hypothesis	Present	461 (82.8)
	Missing	96 (17.2)
Analysis of biomarker	Yes	291 (52.2)
	No	266 (47.8)
Study funding	Private	322 (57.8)
	Private and public	76 (13.6)
	Public	58 (10.4)
	No funding	36 (6.5)
	Not reported	65 (11.7)
Number of center	Multicenter	443 (79.5)
	Single center	43 (7.7)
	Not reported	71 (12.7)
Study results	Positive	245 (44)
	Negative	152 (27.3)
	Non assessable	160 (28.7)
Cut-off point to conclude (*)	Present	246 (56.7)
	Missing	187 (43.1)
	Non assessable	1 (0.2)
Type of primary end point (*)	Rate	372 (80.7)
	Censored data	62 (13.4)
	Incidence	6 (1.3)
	Other	21 (4.6)
Number of primary end point (*)	Unique	461 (91.5)
	Multiple	43 (8.5)
Type of analysis planned	Not reported	287 (51.5)
	Intent to treat	152 (27.3)
	Per-protocol	116 (20.8)
	Other	2 (0.4)
Type of analysis realized	Per-protocol	335 (60.1)
	Intent to treat	215 (38.6)
	Other	7 (1.3)

(*) subgroup descriptive analysis

<https://doi.org/10.1371/journal.pone.0185536.t002>

Table 3. Details of rate of reporting of each item of Key Methodological Score.

Items	Number of trials with item reported (%)
Primary endpoint	504 (90.5)
Sample size justification	371 (66.6)
Evaluable population's definition	222 (39.9)

<https://doi.org/10.1371/journal.pone.0185536.t003>

To our knowledge, only 2 publications assessed reporting in oncology phase II trials with a global score [27,28]. Ottaiano *et al* [28], published an 11-item score, called the Quality Index which is adapted to phase II trials of biotherapy and immunotherapy in oncology. 141 trials published in 5 oncology journals, during a 5-year period, between 1998 and 2002, were analyzed. The Impact Factor of the journal was significantly associated with a score higher than 50/100 (p-value = 0.001). Contrary to Ottaiano *et al*, our analysis with 3 oncology journals found no association with the journal type and reputation, even if the 5-year Impact Factor of the 3 journals selected was well distributed: 16.97 for Journal of Clinical Oncology, 6.89 for Annals of Oncology and 5.31 for British Journal of Cancer. Furthermore, the distribution of trials selected in our analysis were well balanced with 247 trials selected in Journal of Clinical Oncology (Impact Factor >10) and 310 trials selected in Annals of Oncology and British Journal of Cancer (Impact Factor <10).

Grellety *et al* [27] published two global scores for the assessment of reporting quality for oncology phase II trials: the Overall Quality Score and the Key Methodological Score. The first one could be considered, like the CONSORT guidelines, as an aid for the writing of trials, for authors considering the 44 items proposed. The Key Methodological Score, which includes only 3 items, is in fact, from our point of view, the only tool adapted to systematic reporting evaluation in oncology phase II trials. These scores were used by the authors for the analysis of anticancer drugs, radiotherapy and surgery trials. They screened 156 trials published in 8 oncology journals with a 5-year Impact Factor above 4. The period of selection was the year 2011. Rates of reporting of each item of the Key Methodological Score was 68.6%, 77.6% and 33.3% respectively for the reporting of primary endpoint, reporting of the sample size justification and reporting of the definition of an evaluable population. These rates seem to be comparable with our findings for the two last cited items: 66.6% versus 77.6% for sample size justification and 39.9% versus 33.3% for the definition of an evaluable population. Concerning the reporting of primary endpoint, we found a rate higher than Grellety *et al* (90.5% versus 68.6%). Whereas these authors' analysis concerned a larger pool of oncology journals than ours, our analysis concerned a 5-year period which could explain the higher number of screened trials (557 versus 156). A statistical analysis was also conducted by Grellety *et al*. No

Table 4. Factors associated with Key Methodological Score value: Multivariate analysis results.

Variables	Odds ratio ; 95% CI	p-value
Reporting of statistical design (reported versus deduced)	2.22 [1.36–3.65]	<0.001
Number of center (reference: multicenter)		<0.001
Single center	0.25 [0.09–0.74]	0.012
Non assessable data	0.39 [0.19–0.79]	0.009
Type of analysis conducted (reference: intent-to-treat)		<0.001
Per-protocol analysis	0.48 [0.32–0.72]	<0.001
Non assessable data	0 [0–∞]	0.981

<https://doi.org/10.1371/journal.pone.0185536.t004>

factor was associated with Key Methodological Score value, on both univariate and multivariate analysis. However, variables explored by the authors on univariate and multivariate analysis were not as numerous as our variables. Research funding, site of disease and type of treatment were the only similar variables analyzed in the two studies. Impact Factor higher than 10 and registration on clinicaltrials.gov were significantly associated in multivariate analysis to improve the Overall Quality Score.

The quality of reporting was also analyzed on a unique criteria. Reporting of the statistical design appeared to be the most frequently studied criteria [17–23]. Like us, Perrone *et al* [20], showed a significant impact of the number of participating centers on reporting quality, in 145 breast cancer phase II trials, published between 1995 and 1999. Multicenter trials were significantly associated with a higher rate of statistical design reporting (OR = 3.24; 95%CI [1.47–7.15]). In this study, single drug trial, journal Impact Factor and the period of time between the beginning of the study and its publication, were also significantly associated with statistical design reporting.

Some authors analyzed the rate of reporting of one of the three items of the Key Methodological Score. Sample size justification was the object of two descriptive analyses, but no statistical analysis was conducted in order to find factors associated with this criteria [23,29]. Reporting of the primary endpoint in oncology phase II trials was analyzed descriptively [23,30]. Nickolich *et al* [31] showed that the rate of primary endpoint reporting increased significantly with time in 40 small cell, lung cancer phase II trials published in the *Journal of Clinical Oncology* between 1986 and 2010. Indeed, primary endpoint was reported in 10%, 55.6%, 100% respectively for the periods 1986–1996, 1997–2005, 2006–2010 (p-value = 0.01). No significant association was found in our analysis between the year of publication and the Key Methodological Score. We note that the Key Methodological Score publication by Grellety *et al* [27] in January 2014 did not improve the reporting quality of oncology phase II trials during the period January 2014–February 2015 in our analysis. This could be explained by the very short period of time between its publication and the end of our extraction: these new recommendations had yet to be adopted by the scientific community.

One of the biases of our analysis is that phase II trial screenings were performed by mainly one author. Double reading of phase II trials was done only for the studies for which the reviewer hesitated to fill in the data base. Inversely, Ottaiano *et al* [28], and Grellety *et al* [27] assessed each trial with at least two reviewers.

Furthermore, inclusion in the analysis of phase II/III trials can be questionable. Indeed, in Grellety *et al* publication [27], phase II/III trials were excluded from analysis. However, we analyzed only phase II part of phase II/III trials, and we undertake that there is no difference in terms of the need for the quality of reporting between phase II alone and phase II part of phase II/III trials.

The Key Methodological Score, used as a reference score to evaluate the quality of reporting of oncology phase II trials, is debatable. Indeed, with only 3 items, the boundary between “good” and “bad” reporting is difficult to determine. Studies with a score equal to 2 are the most frequent in our analysis (48.5% of trials). On the other hand, the 3 items of the Key Methodological Score, present advantages of reproducibility and facility of use for the reader. The choice of using this score for our analysis was justified by the fact that it was, to our knowledge, the only global score published for systematic analysis of reporting quality for oncology phase II clinical trials. Yet, improvement and refinement of the Key Methodological Score would be of interest, in order to get closer, in its structure, to the other widely spread methodological reporting assessment scores.

This study supports the assumption of non-optimal reporting for the phase II clinical trials in scientific published articles. It also raised the question of the design quality and of the

suitability of the methodological choices in the conduction of these trials. This issue deserves further investigation. Our analysis enriches the poor content in the scientific literature on the subject of the reporting quality of phase II oncology trials, which are a cornerstone of the development of new anticancer therapies. This is to our knowledge, the largest study in terms of the quantity of analyzed trials, and the first to demonstrate a statistically significant association between some study characteristics and the Key Methodological Score value. Because no other global score is available today, the Key Methodological Score must be considered as a reference for the systematic analysis of the quality of reporting in phase II oncology trials, both by authors and readers.

Supporting information

S1 Table. Data set supporting the analyses.
(XLSX)

S2 Table. PRISMA statement check list. Authors are designed by their initials.
(DOC)

Acknowledgments

The authors would like to thank Susan Guillaumond (Activenglish) for translation and editing services.

Author Contributions

Conceptualization: Aurélie Bourmaud.

Data curation: Julien Langrand-Escure, Romain Rivoirard, Mathieu Oriol, Aurélie Bourmaud.

Formal analysis: Mathieu Oriol, Fabien Tinquaut, Aurélie Bourmaud.

Methodology: Mathieu Oriol

Supervision: Frank Chauvin, Nicolas Magné.

Validation: Aurélie Bourmaud.

Visualization: Chloé Rancoule.

Writing – review & editing: Chloé Rancoule, Aurélie Bourmaud.

References

1. Mathoulin-Pélissier C par AK et S. Méthodes biostatistiques appliquées à la recherche clinique en cancérologie n.d. http://www.jle.com/fr/ouvrages/e-docs/methodes_biostatistiques_appliquees_a_la_recherche_clinique_en_cancerologie_290431/ouvrage.phtml (accessed April 25, 2016).
2. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004; 3:711–716. <https://doi.org/10.1038/nrd1470> PMID: 15286737
3. DiMasi JA, Grabowski HG. Economics of New Oncology Drug Development. *J Clin Oncol* 2007; 25:209–16. <https://doi.org/10.1200/JCO.2006.09.0803> PMID: 17210942
4. Parulekar WR, Eisenhauer EA. Novel endpoints and design of early clinical trials. *Ann Oncol* 2002; 13:139–43. <https://doi.org/10.1093/annonc/mdf651> PMID: 12401680
5. Rubinstein LV. Design Issues of Randomized Phase II Trials and a Proposal for Phase II Screening Trials. *J Clin Oncol* 2005; 23:7199–206. <https://doi.org/10.1200/JCO.2005.01.149> PMID: 16192604
6. Piedbois P. [Randomised phase II trials in oncology]. *Bull Cancer (Paris)* 2007; 94:953–6.

7. Gray R. Phase II Clinical Trial Design: Methods in Translational Research from the Genitourinary Committee at the Eastern Cooperative Oncology Group. *Clin Cancer Res* 2006; 12:1966–9. <https://doi.org/10.1158/1078-0432.CCR-05-1136> PMID: 16609005
8. Tuma RS. Examining heterogeneity in phase II trial designs may improve success in phase III. *J Natl Cancer Inst* 2008; 100:164–166. <https://doi.org/10.1093/jnci/djn006> PMID: 18230788
9. Booth CM, Calvert AH, Giaccone G, Lobbezoo MW, Eisenhauer EA, Seymour LK. Design and conduct of phase II studies of targeted anticancer therapy: Recommendations from the task force on methodology for the development of innovative cancer therapies (MDICT). *Eur J Cancer* 2008; 44:25–9. <https://doi.org/10.1016/j.ejca.2007.07.031> PMID: 17845846
10. Adjei AA, Christian M, Ivy P. Novel Designs and End Points for Phase II Clinical Trials. *Clin Cancer Res* 2009; 15:1866–72. <https://doi.org/10.1158/1078-0432.CCR-08-2035> PMID: 19276272
11. Ratain MJ, Humphrey RW, Gordon GB, Fyfe G, Adamson PC, Fleming TR, et al. Recommended changes to oncology clinical trial design: Revolution or evolution? *Eur J Cancer* 2008; 44:8–11. <https://doi.org/10.1016/j.ejca.2007.09.011> PMID: 17981025
12. Ratain MJ, Sargent DJ. Optimising the design of phase II oncology trials: The importance of randomisation. *Eur J Cancer* 2009; 45:275–80. <https://doi.org/10.1016/j.ejca.2008.10.029> PMID: 19059773
13. Penel N, Fournier C, Stéphanie C. Études cliniques de phases 0, 1 et 2 en cancérologie: questions d'actualité. *Bull Cancer (Paris)* 2010; 97:55–63.
14. Cannistra SA. Phase II trials in journal of clinical oncology. *J Clin Oncol* 2009; 27:3073–3076. <https://doi.org/10.1200/JCO.2009.23.1811> PMID: 19451415
15. Seymour L, Ivy SP, Sargent D, Spriggs D, Baker L, Rubinstein L, et al. The Design of Phase II Clinical Trials Testing Cancer Therapeutics: Consensus Recommendations from the Clinical Trial Design Task Force of the National Cancer Institute Investigational Drug Steering Committee. *Clin Cancer Res* 2010; 16:1764–9. <https://doi.org/10.1158/1078-0432.CCR-09-3287> PMID: 20215557
16. Brown SR, Gregory WM, Twelves CJ, Buyse M, Collinson F, Parmar M, et al. Designing phase II trials in cancer: a systematic review and guidance. *Br J Cancer* 2011; 105:194–9. <https://doi.org/10.1038/bjc.2011.235> PMID: 21712822
17. Lee JJ. Randomized Phase II Designs in Cancer Clinical Trials: Current Status and Future Directions. *J Clin Oncol* 2005; 23:4450–7. <https://doi.org/10.1200/JCO.2005.03.197> PMID: 15994154
18. Kramar A, Potvin D, Hill C. Multistage designs for phase II clinical trials: statistical issues in cancer research. *Br J Cancer* 1996; 74:1317. PMID: 8883425
19. Medioni JR, Rycke YD, Asselain B. [New trends in assessment in anticancer treatments by phase II clinical trials]. *Bull Cancer (Paris)* 2000; 87:551–6.
20. Perrone F, Di Maio M, De Maio E, Maione P, Ottaviano A, Pensabene M, et al. Statistical design in phase II clinical trials and its application in breast cancer. *Lancet Oncol* 2003; 4:305–311. PMID: 12732168
21. Thezenas S, Duffour J, Culine S, Kramar A. Five-year change in statistical designs of phase II trials published in leading cancer journals. *Eur J Cancer* 2004; 40:1244–9. <https://doi.org/10.1016/j.ejca.2004.01.008> PMID: 15110889
22. Mariani L, Marubini E. Content and quality of currently published phase II cancer trials. *J Clin Oncol Off J Am Soc Clin Oncol* 2000; 18:429–36.
23. Walter T, Krzyzanowska MK. Quality of Clinical Trials in Gastroenteropancreatic Neuroendocrine Tumours. *Neuroendocrinology* 2012; 96:238–48. <https://doi.org/10.1159/000337662> PMID: 22414794
24. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; 276:637–9. PMID: 8773637
25. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Med Res Methodol* 2001; 1:2. <https://doi.org/10.1186/1471-2288-1-2> PMID: 11336663
26. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *PLoS Med* 2010; 7:e1000251. <https://doi.org/10.1371/journal.pmed.1000251> PMID: 20352064
27. Grellety T, Petit-Moneger A, Diallo A, Mathoulin-Pelissier S, Italiano A. Quality of reporting of phase II trials: a focus on highly ranked oncology journals. *Ann Oncol* 2014; 25:536–41. <https://doi.org/10.1093/annonc/mdt550> PMID: 24419237
28. Ottaviano A, Mollo E, Di Lorenzo G, Pisano C, Di Maio M, Barletta E, et al. Prospective clinical trials of biotherapies in solid tumors: a 5-year survey. *Cancer Immunol Immunother* 2005; 54:44–50. <https://doi.org/10.1007/s00262-004-0567-z> PMID: 15693138

29. El-Maraghi RH, Eisenhauer EA. Review of Phase II Trial Designs Used in Studies of Molecular Targeted Agents: Outcomes and Predictors of Success in Phase III. *J Clin Oncol* 2008; 26:1346–54. <https://doi.org/10.1200/JCO.2007.13.5913> PMID: [18285606](https://pubmed.ncbi.nlm.nih.gov/18285606/)
30. Riechelmann R, Dounaevskaia V, Krzyzanowska M. Quality of reporting primary outcomes in phase II cancer trials. *J Clin Oncol* 2008. <https://doi.org/10.1200/JCO.2008.19.7343> PMID: [18955436](https://pubmed.ncbi.nlm.nih.gov/18955436/)
31. Nickolich M, Babakochi S, Fu P, Dowlati A. Clinical trial design in small cell lung cancer: surrogate end points and statistical evolution. *Clin Lung Cancer* 2014; 15:207–12. <https://doi.org/10.1016/j.clc.2013.12.001> PMID: [24485231](https://pubmed.ncbi.nlm.nih.gov/24485231/)

IV- Discussion

Le fil conducteur du présent travail de thèse a donc consisté à analyser les différents aspects du reporting des publications d'études en oncologie, que ce soit pour les études interventionnelles (essais de phase II, essais de phase III) ou les études observationnelles. De manière plus large, l'objectif des différents travaux réalisés était de mettre en lumière le lien existant entre l'amélioration de la qualité du reporting des publications et l'optimisation de la qualité de la recherche en général. À la suite du recueil des résultats de la première revue de littérature [199], il a été constaté que le reporting des publications d'essais de phase III demeurait un sujet de recherche prédominant dans ce domaine : 52 revues portaient sur ce sujet, soit 90% des articles inclus. L'essai de phase III en oncologie est un essai le plus souvent de grande envergure, multicentrique, où plusieurs centaines à milliers de patients sont inclus. L'enjeu de ces essais est important car, selon les résultats de ces essais, une autorisation de mise sur le marché peut être octroyée ou non. Sur le plan du schéma expérimental, il s'agit toujours d'essais contrôlés randomisés, idéalement en double aveugle. La stratégie consiste ici à minimiser le risque de biais. La transparence des informations (d'ordre méthodologique, d'efficacité, de toxicité notamment) contenues dans les publications d'essais de phase III reste donc un enjeu majeur, dans une optique de soumission de dossier d'autorisation de mise sur le marché pour une nouvelle thérapeutique ou un nouveau protocole de traitement. Afin d'optimiser cette transparence, des premières recommandations de reporting pour les essais cliniques randomisés ont été publiées en 1996 : Le CONSORT [55]. Ces recommandations avaient pour objectif principal d'être un outil d'aide à la rédaction des manuscrits pour les auteurs, afin que le maximum d'informations pertinentes apparaisse dans la publication finale. Du côté des éditeurs de journaux médicaux et des comités de relecture, ces recommandations représentaient un outil de référence afin de vérifier que l'ensemble des informations apparaissent dans les manuscrits soumis avant publication. Depuis 1996, les recommandations CONSORT ont fait l'objet de mise à jour en 2001 [64] et 2010 [65]. Ces recommandations sont désormais soutenues par plus de 400 revues et par plusieurs groupes de rédactions. Le reporting des publications d'essais de phase III en oncologie est donc un sujet qui a été largement abordé dans des études antérieures, et les recommandations de reporting CONSORT sont désormais utilisées par les éditeurs et largement diffusées.

La suite du travail de recherche s'est donc concentré sur le reporting de publications d'études moins analysées jusqu'à présent : Les essais de phase II et les études observationnelles d'oncologie.

Suite à la publication de CONSORT, de nombreuses autres recommandations de reporting ont été publiées : Le REMARK en 2005, pour les études de marqueurs tumoraux et les études médicales pronostiques [86], le STROBE en 2007, pour les études observationnelles [71], le PRISMA, en 2009, pour les revues systématiques de la littérature et les méta-analyses [82] sont quelques exemples de recommandations publiées ces dernières années. Toutefois, il n'existe à ce jour aucune recommandation spécifique de reporting

publiée pour les essais de phase II d'oncologie et on peut s'interroger sur la qualité globale de leur reporting, ainsi que sur l'éventuelle nécessité d'élaborer de telles recommandations.

Il a été démontré dans le travail d'introduction que l'essai de phase II fait l'objet d'une méthodologie très spécifique, aussi bien dans le schéma expérimental utilisé que dans le calcul du nombre de sujets à inclure ou dans le choix du critère de jugement principal. Deux revues systématiques du travail de thèse se sont donc focalisées sur le reporting des articles d'essais de phase II en oncologie [200-201], afin de mettre en lumière d'éventuelles pertes d'informations dans les publications, de décrire les outils actuellement utilisés pour optimiser ce reporting et de rechercher les caractéristiques de ces essais de phase II ayant pu influencer la transparence de leur reporting.

Les premiers résultats du présent travail de thèse ont objectivé que les différentes étapes méthodologiques des essais cliniques de phase II en oncologie sont relativement clairement rapportées dans les publications. Les résultats de la quatrième revue systématique de littérature [200] ont toutefois mis en évidence un reporting faible d'éléments tels que les détails sur le calcul du nombre de sujets nécessaires (66.6% des articles) ou de la définition, dans la partie Méthodes du manuscrit, de la population évaluable (39.9% des articles). De même, le schéma expérimental était clairement rapporté dans 395 publications, soit seulement 70.9% des articles analysés. A l'opposé, la définition du critère de jugement principal était clairement rapportée dans 504 publications, soit 90.5% des articles et le financement de l'étude était rapporté dans 88.3% des cas.

Ces données sont concordantes avec la littérature : Le reporting des éléments justifiant le calcul du nombre de patients à inclure dans l'essai était retrouvé pour 53 à 61% des articles d'essais de phase II, dans 2 études [202, 203] analysées dans la deuxième revue systématique de ce travail de thèse [201]. Concernant le schéma expérimental des études de phase II, ce dernier était correctement rapporté dans moins de 50% des cas, pour 6 publications ayant évalué cet item. Ceci représente une proportion moins importante de reporting adéquat du schéma expérimental que dans la quatrième revue systématique du travail de thèse, mais les articles analysés dans la quatrième revue du travail de thèse avaient été publiés plus récemment (entre 2010 et 2015). Dans l'étude de Nickolich et al [196], la définition du critère de jugement principal était aussi clairement rapportée dans 100% des publications d'essais de phase II publiées entre 2006 et 2010.

Ces résultats constituent un premier argument en faveur de l'élaboration de recommandations de reporting spécifiques aux essais de phase II en oncologie : L'utilisation par les auteurs de recommandations spécifiques pour les essais de phase II, permettrait dans le futur la publication d'articles dont le reporting serait plus complet, plus homogène et plus transparent. Du côté des comités de relecture de journaux médicaux, ces recommandations de reporting pour les essais de phase II pourraient être utilisées comme un outil de vérification de la transparence des différentes actions réalisées au cours de l'étude, avant une éventuelle publication et diffusion des résultats.

La deuxième revue systématique de la littérature a, d'une part, permis de constater qu'il existe peu de données concernant l'évaluation de la qualité du reporting des essais de phase II en oncologie (seulement 13 publications) et d'autre part qu'il n'existe que trois outils utilisés dans les études, ayant pour but d'évaluer, et donc à terme d'optimiser, le reporting de ces essais [193,194]. Ces trois outils ont été décrits dans deux études. Grellety et al [193] avaient élaboré un score global d'évaluation du reporting (Overall Quality Score ou OQS), basé sur 44 éléments à vérifier dans les publications d'essais de phase II en oncologie. Chaque item compris dans la liste couvrait une partie spécifique de l'article (Titre, introduction, méthodes, résultats, discussion) et comptait pour un point s'il était correctement décrit. Sur les 156 articles d'essai de phase II analysés, le score OQS médian était égal à 28 : $OQS_{\text{median}} = 28/44$ (min-max : 9-35). Les mêmes auteurs avaient également utilisé dans cette étude un second outil d'évaluation du reporting intitulé KMS (Key Methodological Score). Ce score prenait en compte 3 items méthodologiques importants, comptant chacun pour un point lorsqu'il était présent dans l'article : La définition claire du critère de jugement principal, la justification claire du nombre de patients inclus dans l'étude et la définition à priori de la population dans laquelle a été évalué le critère de jugement principal et secondaire. Le score KMS médian était égal à 2/3 et 16.1% des études avaient un score KMS maximal de 3/3. Cet outil KMS a été réutilisé lors de la réalisation de la quatrième revue systématique de la littérature [200] : Le score KMS médian était égal à 2/3 (interquartile Q1-Q3 : 1 - 3), avec un score KMS maximal de 3/3 pour 26.2% des 557 articles analysés. Enfin, dans l'étude d'Ottaiano et al [194], les auteurs ont utilisé un troisième score de reporting, nommé index de qualité de reporting, avec une checklist de 11 items, dont le score global était compris entre 0 et 100. L'index de qualité médian des 141 articles analysés était égal à 53/100 (min-max : 18-86), avec une grande hétérogénéité entre les différents essais de phase II.

Il est important de préciser que ces trois outils d'évaluation du reporting, décrits et utilisés dans la littérature, ne sont en aucun cas des outils officiellement validés et utilisés de façon systématique par les différents comités de relecture de journaux médicaux. Ceci constitue donc un deuxième argument pour la conception puis la publication de recommandations officielles de reporting pour les essais de phase II en oncologie. Les auteurs auraient ainsi à leur disposition des recommandations internationales de reporting, applicables pour tous, reproductibles et permettant de limiter au maximum la perte d'information au décours de la publication. Du côté des éditeurs et des comités de relecture, la disponibilité de telles recommandations permettrait de standardiser la procédure de relecture, de réaliser une évaluation objective et reproductible des informations contenues dans le manuscrit par le biais de recommandations validées.

Dans la quatrième revue de la littérature, une analyse univariée puis multivariée avaient été réalisées afin d'identifier des variables présentes dans les articles d'essais de phase II, possiblement associées à un score KMS maximal (KMS=3), versus tous les autres scores du KMS (KMS<3). En analyse multivariée, 3 variables étaient significativement associées à la valeur du score KMS (égal à 3 contre <3) : Le reporting clair versus déduit du design statistique (OR = 2.22; IC 95% [1.36-3.65]; valeur p <0.001), l'essai monocentrique versus multicentrique (OR = 0.25 [0.09 - 0.74]; valeur p = 0.012) et l'analyse menée per protocole

versus selon intention de traiter (OR = 0.48; IC 95% [0.32 -0.72]; valeur p <0.001). En clair, le reporting transparent du design statistique dans une publication d'essai de phase II est un des facteurs associés à un bon reporting combinés des 3 éléments suivants composant le KMS : Définition claire du critère de jugement principal, la justification claire du nombre de patients inclus dans l'étude et la définition à priori de la population dans laquelle a été évalué le critère de jugement principal et secondaire. Cette constatation semble logique : En l'absence de description précise du design de l'essai (essai mono bras, essai randomisé, essai multibras non randomisé), il semble difficile dans la publication de décrire ensuite la justification du nombre de sujets nécessaires ou de définir le critère de jugement principal. Les résultats mettent aussi en évidence un deuxième facteur, le caractère multicentrique de l'essai, comme étant un facteur associé à un score KMS maximal. Il faut toutefois modérer ces résultats, car les essais de phase II d'oncologie actuel sont essentiellement multicentriques, afin d'assurer un recrutement optimal des patients dans les études et d'éviter des biais de sélection géographiques ou ethniques. Les résultats de la quatrième revue de littérature [200] avaient objectivé seulement 43 publications d'essais de phase II monocentrique, soit seulement 7.7% de l'ensemble des publications d'essais analysés. La proportion d'essais monocentrique de phase II en oncologie est donc faible et devrait logiquement diminuer au fil du temps. Enfin le troisième facteur identifié, le reporting d'une analyse en intention de traiter, est en lien étroit avec l'item du KMS définition à priori de la population dans laquelle a été évalué le critère de jugement principal et secondaire : Lorsqu'il est annoncé une analyse en intention de traiter, il est indispensable de compléter cette information en décrivant non seulement la population attendue, incluse mais surtout évaluée dans l'essai. Lorsque l'analyse est réalisée en intention de traiter, la population incluse dans l'étude correspond parfaitement à celle analysée, afin d'éviter le biais d'attrition et de maintenir la comparabilité des groupes. Cela nécessite toutefois d'être vérifié, afin d'évaluer si l'analyse est effectivement en intention de traiter, et non de type per protocole.

4-1 Un plaidoyer pour des recommandations internationales

Dans les publications des essais de phase II en oncologie, certains items de reporting semblent donc être liés : Lorsqu'il est observé un manque de transparence pour un point précis de l'étude dans la publication (souvent un élément d'ordre méthodologique), cela peut engendrer une perte d'information supplémentaire pour un autre élément de l'étude. Cette observation est donc un argument supplémentaire en faveur de l'élaboration de recommandations officielles de reporting pour les essais de phase II en oncologie. La mise à disposition de ces recommandations permettrait aux auteurs d'éviter des pertes d'informations importantes lors de la rédaction des manuscrits, notamment sur des éléments tels que le design statistique ou l'analyse en intention de traiter. Pour les éditeurs et des comités de relecture, l'adhésion à ces recommandations serait une aide importante pour vérifier non seulement la précision générale du reporting des publications, mais surtout de vérifier la transparence d'items d'ordre méthodologique importants dans les manuscrits soumis.

Certains arguments ont donc été mis en avant, en faveur de l'élaboration de recommandations de reporting pour les essais de phase II en oncologie. On peut toutefois se poser la question suivante : Après la publication de telles recommandations de reporting, cela

a-t-il vraiment un impact sur la transparence du reporting des articles d'études publiées ultérieurement ? Plusieurs données d'études confirment l'effet bénéfique de la publication de telles recommandations sur la transparence des publications ultérieures. En premier lieu, la première revue de littérature réalisée dans le travail de thèse avait calculé un score de reporting global, intitulé Overall Quality Score (OQS) pour les 58 revues de littératures incluses [199]. Ce score OQS s'était inspiré des items décrits dans la déclaration PRISMA, publiée en 2009. Il a été observé dans les résultats un score OQS_{moyen} plus élevé dans le groupe des revues de littératures publiées après la parution de PRISMA, par rapport au groupe des revues publiées antérieurement à PRISMA : $OQSR_{moyen\ 2010-2014} = 10$ (Ecart-type :2) contre $OQSR_{moyen\ 1996-2009} = 8$ (Ecart-type :2) avec $p= 0.0053$ (test de Wilcoxon). Concernant le CONSORT, plusieurs études se sont intéressées à l'évolution du reporting après la publication de ces recommandations. Dans l'article de Tardy MP et al [132], il a été évalué la qualité du reporting de publications d'essais cliniques randomisés, réalisés chez les patients adultes atteints d'un gliome de haut grade, sur une période allant de Janvier 1990 à Décembre 2016. Les auteurs avaient élaboré un score de reporting, nommé CONSORT Score (CS), comprenant des items issus de CONSORT. Ce score CS était généralement plus élevé dans les publications récentes : Score CS médian = 13 entre 1990 et 1996 (avant la publication de CONSORT), score CS médian=17 entre 1996 et 2004 et score CS médian=22 après 2005. En analyse multivariée, la publication de l'étude après 2010 était un facteur associé à un score CS plus élevé ($p=0.001$). Chen YP et al [134] ont analysé la qualité du reporting d'articles d'essais cliniques randomisés de radio-chimiothérapie chez des patients atteints de cancer du nasopharynx. Un score de reporting, intitulé Overall Quality Score (OQS), sur 28 points et basé sur la déclaration CONSORT, avait été utilisé pour l'évaluation. Après analyse multivariée, une publication de l'article après 2010 (soit après la publication de la version mise à jour de CONSORT [65]) était un facteur associé à un score OQS plus élevé.

L'ensemble de ces données illustrent donc bien le rôle de ces recommandations : L'élaboration puis la publication de recommandations de reporting constitue donc une action bénéfique, qui améliore à terme la transparence des informations contenues dans les publications et par conséquent la qualité générale de la recherche. Il faut toutefois que les différents journaux médicaux adhèrent et utilisent ces recommandations pour que ce processus devienne efficace.

C'est à l'aune de ces différents résultats qu'on mesure la pertinence et la nécessité de travailler sur des recommandations internationales de reporting pour les essais de phase II d'oncologie. L'élaboration de ce projet pourrait s'inspirer de la procédure mise en œuvre par le passé [51-54] et ayant abouti aux recommandations CONSORT [55] :

La première étape serait tout d'abord de fonder un consortium de recherche comprenant des experts en méthodologie, cliniciens dans le domaine de l'oncologie et éditeurs de journaux ayant déjà travaillé et publié des articles portant sur le sujet du reporting des publications d'essais de phase II. La procédure décrite ci-dessous se base sur la méthode Delphi, qui a l'avantage d'obtenir l'avis justifié d'un panel d'experts, en se basant sur la rétroaction tout en évitant la confrontation entre experts et en préservant leur anonymat [204]. Après une phase préliminaire d'identification des membres potentiels qualifiés et attribution

d'un numéro d'anonymat pour chaque expert, une première invitation par mail serait envoyée, afin de participer à un premier atelier dont l'objectif serait l'élaboration d'une liste d'items pertinents devant impérativement apparaître dans une publication d'essais de phase II en oncologie. Pour les experts ayant répondu positivement à l'invitation, un questionnaire comprenant une liste très exhaustive d'items et de domaines de reporting, pouvant potentiellement apparaître dans les publications, serait envoyée sous forme numérique. Il serait demandé aux experts d'évaluer pour chaque item ou domaine son degré de pertinence : Élément qui semble indispensable dans la publication, élément de moindre importance mais qui apporte des informations complémentaires, élément de reporting plus spécifique à un design d'étude particulier, élément qui ne semble avoir aucun intérêt. Ces domaines de reporting se focaliseraient sur les caractéristiques générales (contexte et rationnel de l'étude, essai mono/multicentrique, taille d'échantillon, consentement éclairé, description des critères d'inclusion/exclusion des patients, description du critère de jugement principal et des critères secondaires, description du type d'intervention réalisée dans l'étude, numéro d'enregistrement de l'étude sur le site Clinicaltrial.gov [205]) et sur les éléments relatifs à la validité interne d'un essai de phase II (plan expérimental utilisé afin de répondre à l'hypothèse initiale, définition des valeurs d'hypothèse nulle et d'hypothèse alternative, définition de la valeur seuil à atteindre, description de la technique de randomisation en cas d'essai de phase II randomisé, description du suivi réalisé lors de l'étude, description des tests statistiques réalisés). Chaque expert aura également la possibilité de répondre au questionnaire par un ou plusieurs commentaires et suggestions.

Au décours de ce premier tour, il sera mis en lumière des points de convergence et de divergence entre experts. Un premier rapport des résultats sera envoyé sous format numérique à l'ensemble des participants, avec d'éventuelles précisions sur la sélection et la définition des différents items de reporting, les experts ayant alors ensuite la possibilité de réviser leurs évaluations. Le consensus des opinions entre experts sera interprété de la manière suivante : Fort (entre 80 et 100%), modéré (entre 60 et 79.9%), faible (entre 50 et 59.9%) et absence de consensus (entre 0 et 49.9%). Sur les points de divergence (convergence inférieure à 80%), un deuxième tour sera ensuite organisé, cette fois-ci sous la forme d'un atelier de travail où chaque membre du consortium sera invité à être présent physiquement, à un lieu et une date fixée à l'avance afin d'optimiser le taux de participation. Les différents points de divergences seront préalablement regroupés en différentes thématiques. L'atelier sera décomposé en sessions de groupes et en sessions plénières. Pour chaque thématique, un petit groupe de travail sera formé, composé exclusivement d'experts invités. Chaque groupe sera composé idéalement d'un clinicien spécialiste, d'un méthodologiste et d'un éditeur de journaux, avec parmi eux un président de session et un rapporteur. Chaque session plénière sera supervisée par un président invité. La méthode Delphi, décrite ci-dessus, sera réutilisée au sein de chaque groupe de travail, dans l'objectif d'aboutir à un consensus intragroupe. Le président et le rapporteur de chaque groupe présentera ensuite les résultats de leurs sessions respectives lors de la session plénière, et ces résultats feront ensuite l'objet de discussions de la part de l'ensemble des participants.

A l'issue de ce deuxième tour, un deuxième rapport des résultats sera envoyé, en mettant en évidence les éventuels points de divergence restants. Un troisième tour sera probablement nécessaire, afin d'obtenir un consensus optimal et pour la rédaction des recommandations.

La méthode Delphi semble être la méthode la plus pertinente, dans le cas présent, car elle aide à consigner les opinions d'experts dans un domaine bien précis, elle ouvre la possibilité de recueillir les opinions à distance via internet, elle limite le nombre de participants pour l'élaboration des recommandations, facilitant ainsi le travail d'identification et de sélection des experts. Cette méthode est également flexible dans sa conception et dans son administration du questionnaire, elle permet d'obtenir avec certitude un consensus à l'issue des questionnaires successifs. La méthode Delphi facilite également la rétroaction contrôlée, qui consiste en une série d'étapes au cours desquelles un sommaire de l'étape précédente est communiqué aux participants, permettant à ces derniers, s'ils le désirent, de réviser leurs jugements antérieurs.

En revanche, cette méthode peut se révéler longue, fastidieuse, coûteuse et intuitive plutôt que rationnelle. Une autre limitation consiste dans la nature contraignante (plusieurs tours d'enquête) de la procédure. Enfin, la méthode Delphi demeure résolument et exclusivement qualitative, puisqu'elle ne prétend pas à l'analyse statistique.

Pour aboutir à la publication de recommandations de qualité concernant le reporting des essais de phase II en oncologie, l'une des difficultés majeures devant être soulevée est la grande hétérogénéité des plans expérimentaux qui peuvent être utilisés dans ces études : Contrairement aux essais de phase III d'oncologie, qui sont majoritairement des essais randomisés contrôlés, si possible en double aveugle, avec analyse de données de survie, les essais de phase II en oncologie peuvent suivre des plans expérimentaux très divers (essais randomisés ou non, comparatifs ou non, designs statistiques en une étape ou multi étapes [25-28] ou approche Bayésienne [32-34]). Cette hétérogénéité doit donc être prise en compte en amont de la procédure d'élaboration de recommandations sur le reporting.

Cette problématique peut être abordée selon une approche d'analyse des données de la littérature existante, afin de mettre en lumière les plans expérimentaux actuellement les plus utilisés dans les essais de phase II d'oncologie, et surtout d'observer s'il existe, au cours de ces dernières années, une tendance vers une modification des designs d'études. Le recueil de telles données permettrait de savoir s'il existe des plans expérimentaux prédominants, qui devront être obligatoirement intégrés dans les recommandations de reporting initiales. Pour les plans expérimentaux moins utilisés, ces derniers pourraient ensuite être intégrés dans des extensions de recommandations, de façon similaire à ce qui a déjà été réalisé pour les essais de phase III et CONSORT [56-58].

Une première étude, réalisée par Ivanova A. et al [206] et publiée en 2016, donne de premières informations sur le sujet : Dans cette revue de littérature, les auteurs ont recherché dans cinq journaux, (Annals of Oncology, Cancer, Clinical Cancer Research, Journal of Clinical Oncology et Lancet Oncology) l'ensemble des essais de phase II d'oncologie, publiés sur trois années bien distinctes (2005, 2010 et 2014). Les auteurs justifient le choix de ces trois

années, en précisant que, lors de cette période de 9 ans étudiée, de nombreux traitements anticancéreux ont été développés, avec de nouveaux mécanismes d'actions ayant abouti à une augmentation importante d'enregistrements de nouveaux médicaments par la Food and Drug Administration américaine. L'objectif principal a consisté à observer l'évolution, sur ces trois années distinctes, de la proportion dite de "succès" des essais de phase II, définie comme la proportion d'essais de phase II ayant rejeté l'hypothèse nulle. Les objectifs secondaires ont consisté à analyser la transparence du reporting des designs statistiques utilisés et à rechercher une éventuelle association entre proportion de "succès" élevée parmi les essais de phase II et type de méthodologie statistique utilisée.

347 publications d'essais de phase II ont été incluses dans l'étude : 141 publiées en 2005, 150 publiées en 2010 et 56 publiées en 2014. Sur l'ensemble des études, il y avait une majorité d'essais multicentriques (255 études soit 73% des cas) avec une tendance plutôt en faveur d'une diminution du nombre d'essais monocentriques entre les années 2005 et 2014 (47 études soit 33% des cas en 2005, 40 études soit 27% des cas en 2010 et 5 études soit 9% des cas en 2014). De plus, il a également été observé une tendance en faveur d'une augmentation du nombre d'essais internationaux en 2014, par rapport aux années 2010 et 2005 (20 études soit 36% des cas en 2014, 31 études soit 21% des cas en 2010 et 20 études soit 14% des cas en 2005). Concernant le critère de jugement principal utilisé dans les études, le taux de réponse ou réponse objective apparaît comme le critère le plus souvent rapporté (191 études soit 55% des cas et pour 2014, 24 études soit 43% des cas). En deuxième position après le taux de réponse, le critère de jugement principal le plus souvent rapporté était la survie globale, la survie sans progression ou le temps jusqu'à progression (82 études, soit 24% des cas). Les auteurs ont toutefois remarqué qu'il y avait une tendance en faveur d'une augmentation de l'utilisation de ces derniers critères en 2014 (23 études soit 41% des cas en 2014, 28 études soit 19% des cas en 2010 et 31 études soit 22% des études en 2005).

Les résultats suivants ont porté sur le design statistique utilisé dans les 347 publications d'essais de phase II. L'essai monobras était le plus souvent rapporté (218 études soit 63% des cas). L'essai de phase II avec au moins deux bras de traitement restait donc minoritaire, mais avec une tendance en faveur d'une augmentation de l'utilisation de ce type de schéma en 2014 (28 études soit 50% des cas en 2014, 51 études soit 34% des cas en 2010 et 50 études soit 35% des cas en 2005). Parmi les essais de phase II multibras, 68% d'entre eux étaient des essais randomisés et 35,6% étaient des essais comparatifs. Concernant le design précis utilisé dans ces essais, le plan expérimental en une étape (95 études, 44% des cas), le plan multi-étape de Simon (87 études, 40% des cas) et le plan multi-étape de Fleming (12 études, 6% des cas) étaient les plus souvent rapportés. Seulement 7 études (3% des cas) avaient utilisé un design bayésien, avec le plus souvent un manque de transparence dans la description du design de l'étude.

En ce qui concerne l'objectif principal de cet article, la proportion d'essais de phase II ayant rejeté l'hypothèse nulle avait augmenté au cours du temps : 38% des cas pour les études publiées en 2005, 56% des cas pour les études publiées en 2014. De façon assez surprenante, les auteurs avaient remarqué que le résultat principal de l'essai de phase II concernant le rejet ou non de l'hypothèse nulle n'était souvent pas clairement rapporté, que ce soit dans l'abstract

de l'étude ou dans le corps de l'article. Il a ensuite réalisé une analyse par régression logistique, afin de rechercher un éventuel lien entre certains facteurs comme le critère de jugement principal choisi ou le nombre de patients inclus, et le rejet de l'hypothèse nulle. Aucune association statistiquement significative n'a été identifiée au décours de cette analyse.

Pour résumer, les résultats de cette première étude ont mis en avant les éléments suivants : Premièrement, le taux de réponse demeure un critère de jugement encore largement utilisé pour les essais de phase II en oncologie. Deuxièmement, les plans expérimentaux les plus souvent rapportés dans les essais de phase II en oncologie sont les plans en une étape et le plan multi-étape de Simon, avec une utilisation encore marginale de design bayésien. Il faut souligner la proportion importante d'articles analysés dans cette étude, dont le reporting du design statistique n'était pas complet. De plus, de nombreuses publications ne précisent pas parmi leurs résultats si l'hypothèse nulle de départ est rejetée ou non. Ces éléments confortent l'idée du besoin d'élaborer des recommandations de reporting pour les essais de phase II en oncologie.

El-Maraghi RH et al ont publié une étude en 2008 [203] qui s'est intéressée plus particulièrement aux designs statistiques utilisés dans les essais de phase II de Thérapies Moléculaires Ciblées (TMC) en oncologie. Les TMC se distinguent des chimiothérapies cytotoxiques ou cytostatiques conventionnelles notamment par leur mode d'action, qui consiste le plus souvent à bloquer une voie de signalisation précise de l'oncogénèse. La problématique, fréquemment retrouvée lors du développement de ces thérapeutiques, est que cette ou ces voies de signalisation sont parfois absentes et/ou parfois résistantes chez une proportion variable de patients traités. Un enjeu important de ce type d'étude consiste ainsi à définir précisément au début de l'étude la sous population de patients, qui a théoriquement la plus grande probabilité de bénéficier du traitement expérimental afin de maximiser les chances de poursuivre le développement du médicament TMC en phase III. Cette identification de sous population nécessite le plus souvent des recherches préalables de biomarqueurs (tumoral ou sanguin) prédictifs de réponse. Les auteurs se sont donc interrogés sur les designs statistiques utilisés dans ce type particulier d'essai de phase II, sur les critères de jugement utilisés et surtout sur une éventuelle relation entre le type de résultat obtenu au décours de l'essai et l'obtention ultérieure d'une autorisation de mise sur le marché pour le médicament en cours de développement.

Dans cette revue de littérature, les auteurs se sont focalisés sur une série de 31TMC, qui avaient déjà servi de base à une revue publiée en 2004, portant sur le design des essais de phase I en oncologie. Tous les agents étudiés avaient terminé la phase I de leur développement depuis longtemps, ce qui a permis aux auteurs d'inclure, le cas échéant, les publications d'essais de phase II de ces mêmes agents. Parmi les critères de sélection des essais de phase II, la revue de littérature a porté sur six types de tumeurs solides : Sein, colorectal, poumon, ovaire, prostate et rein et les articles devaient être publiés avant le 30 Avril 2006. Les essais avec combinaisons de traitements et les essais en cours ont été exclus de l'analyse. Les auteurs ont inclus les publications en texte intégral et les publications d'abstracts.

Une petite proportion de publications reportait dans un seul article les résultats de plusieurs essais (par exemple, analyse de plusieurs TMC pour une même indication dans le cadre d'un essai de phase II randomisé non comparatif, essai de phase II multibras d'une seule TMC mais avec plusieurs niveaux de dose, essai de phase II évaluant l'efficacité d'une seule TMC mais pour plusieurs types de cancers, chacun d'entre eux ayant été analysé séparément). Afin de décrire les résultats de chaque groupe actif de la TMC de manière appropriée, les auteurs ont considéré chaque étude avec n groupes de traitement actifs potentiels dans laquelle il n'était pas prévu de procéder à une comparaison statistique formelle comme étant n essais rapportés dans une même publication. Dans les quelques publications où une randomisation TMC contre traitement de référence ou placebo avait été réalisée, les groupes de l'étude sans TMC ni placebo n'ont pas été considérés comme des essais distincts et le résultat rapporté était celui qui était prévu pour l'étude globale. Les auteurs ont ensuite procédé à une série de calculs, afin de déterminer le taux de réponse et le taux de progression dans chaque essai. Dans un deuxième temps, le taux de réponse global a été calculé pour chaque TMC. Il a ensuite été recherché, pour chaque TMC, si une autorisation de mise sur le marché avait été accordé ultérieurement par la Food and Drug Administration (FDA). La relation entre le taux de réponse global pour chaque TMC, divisé en quatre groupes (0%, >0% à $\leq 10\%$, >10% à $\leq 20\%$, >20%) et la probabilité d'approbation ultérieure du médicament par la FDA a été évaluée par le test de tendance linéaire de Cochran Armitage.

Sur les 31 TMC étudiées, il a été retrouvé pour 19 d'entre elles des publications d'essais de phase II en monothérapie, dans au moins un des six sous types tumoraux d'intérêt. Au total, 65 publications ont été identifiées. Plusieurs publications reportaient des résultats sur plusieurs types de tumeurs ou sur plusieurs types ou dosages de TMC. Après la prise en compte des situations citées ci-dessus, le nombre total d'essais de phase II était de 89. La majorité des publications avaient rapporté une utilisation d'un design d'essai de phase II monobras (51 publications, 78% des cas). Un design d'essai de phase II randomisé avait été rapporté dans les 14 autres publications. Le critère de jugement principal le plus souvent rapporté était la réponse objective ou taux de réponse (51 essais de phase II, soit 57% des cas). 10 essais de phase II avait utilisé un critère de jugement principal contenant à la fois le taux de réponse et la non-progression. 16 essais (18% des cas) avaient utilisé comme critère principal la survie sans progression ou la proportion de patients non progressifs à un moment prédéfini. Des efforts avaient été entrepris dans 18 essais de phase II (20% des cas) afin de restreindre la population de patients incluables à un sous type spécifique d'histologie ou d'altération en biologie moléculaire tumorale. Sur les 51 essais de phase II ayant la réponse objective comme critère de jugement principal, 27 (53%) avaient clairement rapporté les hypothèses initiales, qui avaient ensuite servi de base au choix du design de l'essai et au calcul du nombre de sujet nécessaires. 76 essais de phase II avaient rapporté des résultats concernant la réponse objective (en critère de jugement principal ou secondaire). Au moins une réponse objective était retrouvée dans 38 essais de phase II (50% des cas). Le taux de réponse à la TMC était supérieur à 10% dans 19 essais de phase II (25% des cas) et supérieur à 20% dans 8 essais de phase II (10.5% des cas). Plusieurs TMC avaient un taux de non-progression supérieur ou égal à 50% (sorafenib, cetuximab, temsirolimus, gefitinib, trastuzumab). 7 TMC sur les 19 analysées ont finalement abouti à une autorisation de mise sur

le marché par la FDA. Les taux de réponse globaux les plus élevés étaient prédictifs de l'approbation ultérieure de la TMC par la FDA, dans les types de tumeurs examinés ($p = 0.005$).

Nous pouvons relever les informations suivantes de cette deuxième étude portant sur le design expérimental des essais de phase II : Tout d'abord, les essais de phase II portant sur les TMC utilisent majoritairement un design d'essai monobras (78% des cas), ce qui représente une proportion plus élevée que dans les essais de phase II publiés en 2005 et inclus dans l'étude d'Ivanova A. et al (65% des cas) [206]. Le critère de jugement principal le plus souvent rapporté était le taux de réponse ou réponse objective (57% des cas), ces données sont cohérentes avec celles objectivées précédemment dans l'étude d'Ivanova A. et al (55% des cas) [206]. Ce critère principal reste particulièrement intéressant dans les essais de phase II de TMC, car les taux de réponse les plus élevés semblent être prédictifs d'une autorisation de mise sur le marché ultérieure. Le taux de non-progression est une donnée complémentaire importante dans ce contexte, un fort taux de non-progression peut être observé pour certaines TMC, majoritairement utilisées dans le traitement du carcinome rénal. Le taux de non-progression peut être envisagé comme un critère de jugement principal valable dans un essai de phase II évaluant une nouvelle TMC dans une indication de traitement de carcinome rénal. Enfin, le reporting d'un essai de phase II portant sur une TMC en oncologie doit clairement préciser si la sélection des patients à l'inclusion s'est basée sur la recherche d'éventuels biomarqueurs, tumoraux ou sanguins, préalablement validés, et potentiellement prédictifs d'une réponse au traitement.

L'immunothérapie, avec en première ligne les inhibiteurs de points de contrôle immunitaires anti PD-1 anti PD-L1 et les anti CTLA-4, représente une classe thérapeutique récente en oncologie, et particulièrement prometteuse. Ces nouveaux traitements ont déjà bouleversé sur ces dernières années les stratégies thérapeutiques, voire même le pronostic, de certains cancer (par exemple : Le mélanome métastatique, le cancer bronchique non à petites cellules localement avancé et métastatique, le cancer épidermoïde de la tête et du cou métastatique, le cancer rénal et urothélial métastatique). Cependant, le mécanisme d'action des agents d'immuno-oncologie diffère grandement de celui des agents cytotoxiques/cytostatiques et des agents de type thérapie moléculaire ciblée : L'objectif de ces médicaments consiste à provoquer une réponse immunitaire innée dirigée contre la tumeur. Ce mode d'action unique peut entraîner un bénéfice clinique chez le patient, avec des mesures de réponses à l'imagerie très diverses, allant de la classique réponse objective avec réduction conséquente de la taille tumorale, à la situation de pseudoprogression initiale et transitoire suivi d'une seconde phase de réponse partielle prolongée ou de maladie stable prolongée. De plus, par comparaison avec les agents cytotoxiques, les agents d'immunothérapie prennent plus de temps (en moyenne, 3 mois) pour apporter un bénéfice clinique. L'ensemble de ces caractéristiques de l'immunothérapie ont rendu les designs conventionnels d'essais de phase II inefficaces et dysfonctionnels (par exemple, un essai de phase II d'immunothérapie utilisant un plan de Simon avec le taux de réponse comme critère principal ne semble pas adapté). Des données de la littérature ont en effet démontré que l'effet retardé de ces traitements d'immunothérapie entraînait une perte de puissance conséquente lorsque les designs

conventionnels d'essais de phase II étaient utilisés [207]. De plus, comme les réponses antitumorales à l'immunothérapie nécessitent de longues périodes d'évaluation afin d'être scorées, ceci entraîne de grandes difficultés logistiques. De nouveaux designs expérimentaux, plus adaptés aux traitements d'immunothérapie, ont donc été élaborés. Certains de ces designs permettent de réaliser de multiples analyses intermédiaires, voir même un monitoring continu des données issues de l'essai de phase II [207-209].

Lin R et al [210] ont publié en 2019 un article décrivant un design expérimental novateur, spécifiquement adapté aux essais de phase II d'immunothérapie en oncologie. Ce design, intitulé design Bayésien optimal de temps jusqu'à événement (TOP design) pour les essais de phase II, s'est inspiré en partie du design Bayésien optimal pour les essais de phase II (BOP2 design) publié en 2017 [211]. Les auteurs ont décrit le TOP design comme efficient, dans la mesure où il permettrait de prendre des décisions provisoires et en temps de réel de poursuite/arrêt de l'essai de phase II, en la survenue de réponse tardive liée à l'immunothérapie. Le TOP design utiliserait l'ensemble des données de l'essai disponibles et maximiserait la puissance statistique afin de détecter les traitements considérés comme efficaces. Ce type de design est considéré par les auteurs comme flexible dans le nombre d'analyses intermédiaires à réaliser, et serait capable de gérer des critères de jugement simples et plus complexes, dans un cadre uniforme. Les auteurs ont décrit dans leur article des études de simulation, afin d'évaluer les caractéristiques de fonctionnement du design TOP, et de comparer ses performances à trois autres designs pré-existants : Design en deux étapes de Simon [27], le design BOP2 [211], le design Bayésien de monitoring de la futilité [212].

Les principaux résultats ont mis en évidence que le TOP design permettait de raccourcir de 4 à 10 mois la durée d'un essai de phase II, par comparaison aux autres designs Bayésiens. Le TOP design améliorerait également la puissance pour détecter des traitements efficaces, avec des valeurs voisines de 90% et des erreurs de type I bien contrôlées. Les auteurs ont toutefois souligné quelques limitations du TOP design : Il n'a été conçu que pour les essais de phase II monobras et ne permettrait pas actuellement de gérer les critères de jugement de type temps jusqu'à progression ou temps jusqu'au décès du patient. Des efforts de recherches restent donc nécessaires pour que les design Bayésiens puissent prendre en charge les cas plus complexes d'essais de phase II randomisés d'immunothérapie, avec des données de survie ou un critère composite comme critère de jugement principal.

Ces données doivent être confirmées, mais méritent d'être sérieusement prises en compte en amont de l'élaboration de recommandations de reporting pour les essais de phase II en oncologie. Le design expérimental Bayésien semble particulièrement adapté pour les essais de phase II d'immunothérapie en oncologie. Le développement de nouveaux agents d'immuno-oncologie a fortement augmenté depuis la dernière décennie, et les futures recommandations de reporting d'essais de phase II doivent obligatoirement intégrer ceci afin que son contenu ne devienne pas rapidement obsolète. Si les premières données sur l'intérêt des designs Bayésiens dans l'évaluation de l'efficacité des immunothérapies se confirment, il y aurait un intérêt d'élaborer une extension de recommandations, dédiée spécifiquement au reporting des essais de phase II d'immunothérapie en oncologie.

L'analyse de ces différentes données de la littérature ont mis en évidence une utilisation encore très large, pour les essais de phase II en oncologie, des designs d'essais monobras, en une étape ou en deux étapes de Simon, avec le taux de réponse comme critère de jugement principal. Les données de survies sont toutefois de plus en plus utilisées, tout comme la randomisation, et devront impérativement être prise en compte dans les futures recommandations de reporting. Le cas des agents d'immunothérapie semble être particulier, notamment du fait de leurs mécanismes d'action avec efficacité souvent retardée. Dans cette dernière situation, les design Bayésiens semblent plus adaptés, afin d'assurer une multiplicité d'analyses intermédiaires en temps réel. La recherche de design statistique optimal pour les essais de phase II d'immunothérapie, voire de chimio-immunothérapie, demeure clairement un objectif scientifique prioritaire, et des données supplémentaires demeurent nécessaires.

4-2 Quel lien entre qualité du reporting et efficacité d'un médicament ?

Entre 2012 et 2014, la probabilité pour un médicament innovant de passer de la phase I à la mise sur le marché était de 9,6% quel que soit le domaine dans lequel ce médicament était étudié. Des données précisent que ce chiffre tend à être diminué par les taux inférieurs obtenus par les médicaments anticancéreux, sans lesquelles le pourcentage de chance d'enregistrement serait de 11,9% [213]. Concernant les essais de phase III, l'oncologie présente un taux d'attrition moyen estimé comme étant de 11.3 à 14% plus haut par rapport aux taux des autres spécialités [214-215]. Ces données importantes ont abouti aux questions et réflexions suivantes : Ne semble-t-il pas nécessaire d'évaluer plus précisément la transparence du reporting des essais cliniques de phase II en oncologie, et notamment des éléments d'ordre méthodologique, avant de valider la poursuite du développement du médicament dans un essai de phase III de plus grande échelle et nettement plus coûteux ? Existe-il un lien entre la transparence du reporting des essais de phase II et l'obtention de résultats positifs, statistiquement significatifs, et cliniquement pertinents dans l'essai de phase III réalisé ultérieurement et dans la même indication ? Quel est actuellement le taux d'attrition des essais de phase II en oncologie ?

Une cinquième revue systématique de la littérature a donc été menée, afin de répondre à l'hypothèse d'un éventuel lien entre les principales caractéristiques d'essais de phase II en oncologie contenues dans les publications (notamment la clarté du reporting) et la positivité des résultats des essais de phase III réalisés ultérieurement. L'étude est désormais terminée, mais n'a pas encore été finalisée ni soumise en vue d'une publication. Seuls les détails de la partie Matériels et Méthodes seront donc rapportés de façon exhaustive ci-dessous.

Dans la revue de la littérature, la sélection des études s'est tout d'abord basée préalablement sur le choix de trois journaux médicaux. Une liste de journaux médicaux a tout d'abord été élaborée. Les journaux ont été sélectionnés sur les critères suivants : Journal non spécialisé dans un type tumoral, publiant des essais de phase II d'oncologie. Cette liste a ensuite été divisée en trois groupes : Un premier groupe de journaux généralistes avec Impact Factor (IF) supérieur à 6, un deuxième groupe de journaux spécialisés en oncologie à haut IF (supérieur à 6) et un troisième groupe de journaux spécialisés en oncologie à IF moyen

(environ égal à 4). La sélection des journaux a été décidée après une concertation multidisciplinaire, comprenant un méthodologiste (dr A Bourmaud), deux cliniciens spécialisés en oncologie (dr Rivoirard et dr Langrand-Escure) et un statisticien (F Tinquaut).

La liste des 19 journaux sélectionnés était la suivante : Le premier groupe de journaux généralistes à IF>6 comprenait : *New England Journal of Medicine*, *Lancet*, *Journal of the American Medical Association*, *Annals of Internal Medicine*, *PlosMed*, *British Medical Journal*, *JAMA internal Medicine*. Le deuxième groupe de journaux d'oncologie à IF>6 était composé de : *Cancer Research*, *Journal of Clinical Oncology*, *Lancet Oncology*, *Journal of National Cancer Institute*, *CA : A Cancer Journal for Clinicians*, *Annals of Oncology*. Enfin, le troisième groupe de journaux d'oncologie à IF environ égal à 4 était constitué de : *British Journal of Cancer*, *European Journal of Cancer*, *The Oncologist*, *Cancer*, *International Journal of Cancer*, *Journal of the National Comprehensive Cancer Network*. Parmi ces 19 journaux, un statisticien (F Tinquaut) a ensuite procédé à une sélection aléatoire : Dans chaque groupe un numéro, compris entre 1 et le nombre total de journaux compris dans chaque groupe, a été attribué à chaque journal. Un numéro a ensuite été tiré au sort dans chaque groupe (en suivant une distribution discrète uniforme entre 1 et le nombre maximal de journaux dans chaque groupe). Chaque journal correspondant à chaque numéro tiré au sort a ensuite été sélectionné dans l'étude. Les trois journaux sélectionnés étaient *British Medical Journal*, *Annals of Oncology* et *British Journal of Cancer*. Les archives des trois journaux ont ensuite été explorées afin d'identifier l'ensemble des publications originales d'essais de phase II d'oncologie. La dernière recherche a été réalisée en Mars 2018.

Les critères d'inclusions pour les publications d'essais de phase II d'oncologie identifiés étaient les suivants : Essais de phase II ou essais de phase II/III, publiés dans un des trois journaux sus-cités, entre Mars 2010 et Février 2015. Les critères d'exclusion dans la revue de littérature étaient les suivants : Les essais de phase I/II, les réanalyses de sous groupes d'essais de phase II ou de phase II/III, analyses poolées de plusieurs essais de phase II ou de phase II/III, publications ayant rapporté uniquement les résultats de la partie phase III des essais de phase II/III. Les essais de phase II et de phase II/III ayant utilisé un design Bayésien ont également été exclus de l'analyse.

La sélection des études a été réalisée séparément et de façon indépendante par deux des auteurs (J Langrand Escure et M Oriol) et ont été ensuite comparées. Un premier auteur (J Langrand Escure) a sélectionné chaque publication d'essai de phases II et de phase II/III, au cours de la période concernée, manuellement dans les archives informatisées de chaque journal. Les mots clés suivants avaient été recherchés : « phase II » ou « phase 2 » ou « phase II / III » ou « phase 2/3 » présents dans le titre et / ou dans l'abstract et / ou dans le dernier paragraphe de la partie Introduction et / ou de la section Méthodes de l'article. Le deuxième auteur (M Oriol) a procédé à la sélection des publications avec la requête électronique suivante dans la base de données MEDLINE : (((("Br J Cancer. "[Journal] OR " J Clin Oncol. "[Journal] OR " Ann Oncol. "[Journal]) AND ("2010/03" [PDAT] : "2015/02" [PDAT])) ET Clinical Trial, Phase II [ptyp]). Toutes les publications identifiées à la suite de la requête électronique ont ensuite été sélectionnées manuellement par cet auteur afin de ne garder que les essais de phase II et de phase II / III publiés dans les 3 journaux sélectionnés. Les

publications communes entre les deux modes de sélections ont été incluses dans l'analyse. Les publications qui n'étaient pas communes aux deux méthodes de sélections ont été lues par un troisième auteur (A Bourmaud) afin de décider, si ces études pouvaient être incluses dans l'analyse.

Chaque publication sélectionnée a ensuite été analysée par un auteur (J Langrand Escure) avec recueil de multiples variables. Les articles pour lesquels l'auteur a eu du mal à définir l'une des variables à recueillir ont été lues par un deuxième auteur (R Rivoirard) afin de prendre une décision. Les variables suivantes ont été recueillies dans chaque article : Année de publication de l'article, type de traitement évalué (chimiothérapie, hormonothérapie, radiothérapie, immunothérapie), évaluation d'une association médicamenteuse ou non, type de chimiothérapie évaluée en cas d'essai de chimiothérapie (thérapie moléculaire ciblée / thérapie cytotoxique ou cytostatique / anticorps monoclonal / combinaison), site tumoral étudié, analyse de biomarqueurs réalisée (oui / non), design statistique utilisé (randomisé / non randomisé), reporting clair du design statistique (oui / non), essai comparatif ou non en cas d'essai multibras, type de plan statistique utilisé, nombre d'étapes prévues dans l'essai (1,2, ≥ 3), ratio de randomisation pour les essais randomisés (1:1 / autre / non rapporté), valeurs prédéfinies de l'erreur de type I et de type II, reporting de l'hypothèse a priori (rapportée / non déclarée), description de la valeur seuil pour conclure pour les essais non comparatifs (rapporté / non rapporté), pourcentage prévu de perdus de vue ($> 10\%$ / $< 10\%$ / non rapporté), type de financement (public / privé / public et privé / non financé / non déclaré), analyse(s) intermédiaire(s) prévu(es) (oui / non / non rapporté), arrêt précoce de l'essai sur résultats de l'analyse intermédiaire (oui / non / non rapporté), nombre de centres (monocentrique / multicentrique / non rapporté), résultats de l'essai de phase II (positif / négatif), type de critère principal (taux de réponse objective / données censurées / autre), nombre de patients prévus, inclus et analysés, type d'analyse prévue et réalisée, développement d'un essai de phase III ultérieur (oui / non), résultats de l'essai de phase III ultérieur (positifs / négatifs / essai en cours), obtention d'une autorisation de mise sur le marché ou d'une extension d'indication concernant le protocole de traitement ou le médicament initialement évalué dans l'essai de phase II (oui / non / non spécifié).

Le score KMS, préalablement décrit [200], a été utilisé afin d'évaluer la qualité globale de reporting de chaque article d'essai de phase II. Les trois éléments suivants ont systématiquement été recueillis pour chaque article : Reporting du critère de jugement principal de l'étude (oui/non), reporting des informations nécessaires à la réalisation du calcul du nombre de sujets à traiter dans l'étude (oui/non), reporting de la définition de la population évaluable dans l'étude (oui/non). Le critère de jugement principal était considéré comme bien rapporté s'il était détaillé dans la partie Introduction ou Méthodes de l'article. Quand le critère de jugement principal était présent uniquement dans l'abstract de l'article, il a été considéré comme non rapporté. La justification du calcul de la taille d'échantillon dans l'étude nécessitait la définition précise de l'hypothèse a priori et les valeurs des erreurs de type I et de type II. La définition de la population évaluable a été recherchée pour le critère de jugement principal et considérée comme bien rapportée si elle correspondait bien à ce qui avait été détaillé dans la partie Méthodes de l'article. Un point a été attribué à chacun des trois

éléments lorsqu'il avait été bien rapporté dans la publication. Le score KMS minimal était donc de 0 et le score maximal était de 3.

Pour chaque article d'essai de phase II, un auteur (R Rivoirard) a ensuite systématiquement vérifié si des essais ultérieurs de phase III avaient été réalisés, pour le même traitement et dans la même indication. Les recherches ont été effectuées sur la base de données MEDLINE et sur le site Web ClinicalTrials.gov [205]. Pour le sous groupe d'articles d'essais de phase II ayant ensuite abouti à un essai de phase III, des données ont ensuite recueillies sur la conclusion de l'essai de phase III (essai positif/essai négatif/essai en cours). Enfin, des données sur les autorisations de mise sur le marché ultérieures ont été collectées par le même auteur, sur la base de données HAS pour la France et sur la base de données de la FDA pour les États-Unis d'Amérique.

Une analyse descriptive a d'abord été réalisée. Les variables qualitatives ont été décrites à l'aide de taux et les variables quantitatives en utilisant la valeur moyenne avec l'écart type associé et la valeur médiane avec interquartile associé. Le score KMS a été décrit en termes de taux pour chaque score (0, 1, 2 ou 3) et en termes de score médian avec interquartile et valeurs minimales et maximales associées. Des premières analyses univariée et multivariée, ont été réalisées afin de rechercher d'éventuelles caractéristiques, présentes dans les articles d'essais de phase II d'oncologie, liées à la réalisation ultérieure d'un essai de phase III. Les variables suivantes ont été incluses dans l'analyse univariée : Nom du journal, année de publication, site tumoral, design statistique utilisé, reporting du design statistique, combinaison de médicaments évaluée, analyse non comparative ou comparative, type de ratio de randomisation, analyse de biomarqueur, reporting d'une valeur seuil pour conclure, type de financement, arrêt précoce sur analyse intermédiaire, nombre de centres, résultats de l'essai de phase II, type de critère de jugement principal, nombre de patients initialement prévus dans l'étude, comparaison entre le nombre de patients prévus / inclus et prévus/ analysés, type d'analyse initialement prévue et réellement réalisée, valeur du score KMS. Une deuxième analyse univariée et multivariée a été réalisée pour le sous-groupe d'essais de phase III réalisés, afin de rechercher d'éventuelles variables explicatives aux résultats positifs d'un essai de phase III ultérieur. Les mêmes variables que celles décrites ci-dessus ont été incluses dans le modèle univarié. Une régression logistique a été réalisée pour obtenir pour chaque variable un odd ratio avec un Intervalle de confiance (IC) à 95%. Les variables identifiées avec une valeur du petit p inférieure à 0,2 sur l'analyse univariée, et non corrélés entre elles, ont été incluses dans le modèle multivarié et analysées par une nouvelle régression logistique. Les variables avec une valeur de petit p inférieure à 0,05 après analyse multivariée ont été considérées comme significativement associées à la réalisation d'un essai ultérieur de phase III ou associées à des résultats de phase III positifs. Toutes les analyses statistiques ont été effectuées avec le logiciel R 3.1.1.

Les résultats de la cinquième revue de littérature seront prochainement soumis pour publication.

La première revue systématique de la littérature avait mis en évidence le manque de données sur le sujet de l'évaluation du reporting des études observationnelles en oncologie

[199]. Une seule étude s'était intéressée spécifiquement à ce sujet : Dans la revue de Papathanasiou AA et al [150], le reporting de 244 études observationnelles d'oncologie avait été analysé. La qualité du reporting global de ces études avait été considérée comme satisfaisante, avec toutefois certaines informations moins rapportées, comme les détails sur l'appariement dans les études cas-témoin, l'information sur les données manquantes et la publication de diagramme de flux. Ce manque de données concernant les études observationnelles en oncologie a servi de point de départ pour la réalisation du troisième travail de thèse [216]. Cette revue de la littérature a analysé le reporting de 698 articles d'études observationnelles (84.5% des articles analysés dans la revue). Le reporting de la description des variables d'intérêt, dans la partie Méthodes des articles a été considéré comme parfait (reporting transparent dans 100% des articles). Par contre, lorsqu'il a été analysé la concordance de reporting, entre la partie Méthodes et Résultats, pour le test statistique principal utilisé dans ces études, le taux de concordance parfaite n'était que de 42.2%. Il est important de souligner que le reporting précis du test statistique utilisé n'est demandé que pour la partie Méthodes dans STROBE. Concernant la partie Résultats, il n'y a donc aucune recommandation pour les auteurs de rapporter le test statistique finalement réalisé, qui peut différer dans certains cas du test statistique prévu initialement. Ces observations illustrent l'importance de mettre à jour et d'actualiser régulièrement les différentes recommandations de reporting, ce qui a déjà été fait pour CONSORT en 2001[64] et en 2010[65].

Il a été remarqué précédemment que l'élaboration et la publication de recommandations de reporting sont des actions efficaces sur l'optimisation du reporting des articles d'études, uniquement si les éditeurs de journaux médicaux adhèrent à ces recommandations. En 2018, Sharp MK. Et al avaient publié une étude dont l'objectif principal était d'estimer la prévalence et la typologie de l'adhésion à la déclaration STROBE des journaux médicaux [96]. Sur 257 journaux différents publiant des études observationnelles, 139 (54% des cas) ne mentionnaient nulle part la déclaration STROBE dans les instructions aux auteurs, tandis que seulement 12 journaux (5% des cas) exigeaient que les items de STROBE soient respectés au moment de la soumission. Dans cette étude, l'adhésion à d'un journal à la déclaration STROBE était associée à son adhésion au CONSORT et au PRISMA. D'autres études ont signalé un problème d'adhésion des journaux médicaux aux différentes recommandations de reporting : le CONSORT pour les essais cliniques randomisés [118], le STRICTA pour les essais d'acupuncture [181]. Un plus grand nombre de journaux médicaux devraient adhérer aux recommandations de reporting et, plus important encore, il devrait être fait plus d'efforts pour garantir leur adhésion.

Finalement, il est important de reconnaître qu'un « mauvais » reporting d'un article d'étude n'est pas forcément synonyme d'une méthodologie inadaptée. Par exemple, la quatrième revue de littérature a ainsi démontré que le score KMS des articles d'essais de phase II analysés était inférieur ou égal à 1 pour 25% des études, mais que le choix du critère de jugement principal demeurerait pertinent dans la majorité des cas. Transparence du reporting et qualité de la méthodologie d'une étude ne sont donc pas des termes équivalents, et il ne faut pas juger de la qualité globale d'une étude uniquement sur la transparence du reporting de l'article publié.

V- Conclusion

En conclusion, ce travail de thèse a permis de mettre en évidence la nécessité de concevoir des recommandations spécifiques sur le reporting des articles d'essais de phase II en oncologie. Cette action ne peut être efficace que si, d'une part, les auteurs des articles se réfèrent à ces recommandations de reporting lors de la rédaction de l'article et si, d'autre part, les journaux médicaux adhèrent et utilisent activement ces recommandations. Un exemple pratique d'élaboration de recommandations pour les essais de phase II en oncologie a été proposé dans ce manuscrit. Plusieurs études montrent une adhésion inconstante des journaux médicaux aux différentes recommandations existantes (CONSORT, STROBE, PRISMA, etc) et des efforts doivent être réalisés pour augmenter cette adhésion. Le reporting des études observationnelles est un sujet encore peu étudié et des données supplémentaires et actualisées sont nécessaires afin d'apprécier l'évolution de la transparence du reporting pour ces études. Des études sont aussi attendues pour identifier le lien entre qualité du reporting et qualité de la méthodologie utilisée, validité des résultats. Enfin, il est important d'actualiser régulièrement les recommandations de reporting au cours du temps.

VI- Bibliographie

1. Organisation Mondiale de la Santé (OMS) –principaux faits-1^{er} Février 2018. Accessible à l'URL: <http://www.who.int/fr/news-room/fact-sheets/detail/cancer-2>.
2. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016 Oct; 388 (10053):1659-1724.
3. Global Initiative for Cancer Registry Development. International Lyon: International Agency for Research on Cancer
4. Stewart BW, Wild CP, editors. World cancer report 2014 Lyon: International Agency for Research on Cancer; 2014
5. European Cancer Information System (ECIS) – Incidence and mortality historical data. Accessible à l'URL: <https://ecis.jrc.ec.europa.eu/explorer.php>
6. Projection de l'incidence et de la mortalité par cancer en France métropolitaine en 2017-Rapport technique. Accessible à l'URL: <http://www.e-cancer.fr/Professionnels-de-sante/Les-chiffres-du-cancer-en-France/Epidemiologie-des-cancers/Donnees-globales>
7. Principales causes de décès et de morbidité - L'état de santé de la population en France Rapport 2017. Accessible à l'URL: http://drees.solidarites-sante.gouv.fr/IMG/pdf/esp2017_5_principales_causes_de_deces_et_de_morbidite.pdf
8. Plan cancer 2014-2019-Deuxième édition. Page 48, Action5.2. Accessible à l'URL: <http://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Plan-Cancer-2014-2019>
9. Misset JL. Histoire de l'oncologie médicale. Bull Cancer 2013 ; 100 : 403-6.
10. Mathoulin-Pélessier C par AK et S. Méthodes biostatistiques appliquées à la recherche clinique en cancérologie n.d. : http://www.jle.com/fr/ouvrages/e-docs/methodes_biostatistiques_appliquees_a_la_recherche_clinique_en_cancerologie_290431/ouvrage.phtml

11. Sheiner LB. Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* 1997;61:275-91.
12. Sandler A. Clinical experience with the HER1/EGFR tyrosine kinase inhibitor erlotinib. *Oncology* 2003; 17:17-22.
13. Simon R, Freidlin B, Rubinstein L et al. Accelerated titration designs for phase I clinical trials in oncology. *J Natl Cancer Inst* 1997; 89:1138-47.
14. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 1990; 46:33-48.
15. Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. *Stat Med* 1995;14:1149-61.
16. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* 2000;56:1177-82.
17. Thall PF, Estey EH, Sung HG. A new statistical method for dose-finding based on efficacy and toxicity in early phase clinical trials. *Invest New Drugs* 1999; 17:155-67.
18. Braun TM. The bivariate continual reassessment method; extending the CRM to phase I trials of two competing outcomes. *Control Clin Trials* 2002; 23:240-56.
19. Dees EC, Whitfield LR, Grove WR et al. A phase I and pharmacologic evaluation of the DNA intercalator CI-958 in patients with advanced solid tumors. *Clin Cancer Res* 2000; 6:3885-94.
20. Van Kesteren C, Mathot RA, Beijnen JH, Schellens JH. Pharmacokinetic-pharmacodynamic guided trial design in oncology. *Invest New Drugs* 2003; 21:225-41.
21. Ploin DY, Tranchand B, Guastalla JP et al. Pharmacokinetically guided dosing for intravenous melphalan: a pilot study in patients with advanced ovarian adenocarcinoma. *Eur J Cancer* 1992; 28A:1311-5.
22. Korn EL, Simon R. Using the tolerable-dose diagram in the design of phase I combination chemotherapy trials. *J Clin Oncol* 1993; 11:794-801.
23. Morita S, Toi M, Kobayashi T et al. Application of a continual reassessment method to a phase I clinical trial of capecitabine in combination with cyclophosphamide and epirubicin (CEX) for inoperable or recurrent breast cancer. *Jpn J Clin Oncol* 2004; 34:104-6.

24. Diniz MA, Quanlin-Li, Tighiouart M. Dose Finding for Drug Combination in Early Cancer Phase I Trials using Conditional Continual Reassessment Method. *J Biom Biostat.* 2017; 8(6). pii: 381.
25. Fleming TR. One-sample multiple testing procedure for phase II clinical trials. *Biometrics* 1982; 38:143-51.
26. Gehan EA. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *J Chronic Dis* 1961; 13:346-53.
27. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 1989;10:1-10.
28. Ensign LG, Gehan EA, Kamen DS, Thall PF. An optimal three-stage design for phase II clinical trials. *Stat Med* 1994; 13:1727-36.
29. Lee JJ, Feng L. Randomized phase II designs in cancer clinical trials: current status and future directions. *J Clin Oncol* 2005; 23:4450-7.
30. Wieand HS. Randomized phase II trials: what does randomization gain? *J Clin Oncol* 2005;23:1794-5.
31. Conaway MR, Petroni GR. Bivariate sequential designs for phase II trials. *Biometrics* 1995;51:656-64.
32. Estey EH, Thall PF. New designs for phase 2 clinical trials. *Blood* 2003;102:442-8.
33. Berry DA. Bayesian clinical trials. *Nature Reviews Drug Discovery* 2006;5:27-36.
34. Thall PF, Wathen JK. Practical Bayesian adaptive randomisation in clinical trials. *Eur J Cancer* 2007; 43:859-66.
35. Inoue LY, Thall PF, Berry DA. Seamlessly expanding a randomized phase II trial to phase III. *Biometrics* 2002; 58:823-31.
36. Storer BE. A sequential phase II/III trial for binary outcomes. *Stat Med* 1990;9:229-35.
37. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *J Am Med Assoc* 1995; 273:408-12.
38. Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. *Consolidated Standards of Reporting Trials. J Am Med Assoc* 1998;279:1489-91
39. Kjaergard LL, Villumsen J, Gluud C. Quality of randomised clinical trials affects estimates of intervention efficacy. Rome, Italy: 7th Cochrane Colloquium; 1999.

40. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *Br Med J* 2001; 323:42-6.
41. Veldhuyzen van Zanten SJ, Cleary C, Talley NJ, Peterson TC, Nyren O, Bradley LA, et al. Drug treatment of functional dyspepsia: a systematic analysis of trial methodology with recommendations for design of future trials. *Am J Gastroenterol* 1996; 91:660-73.
42. Talley NJ, Owen BK, Boyce P, Paterson K. Psychological treatments for irritable bowel syndrome: a critique of controlled treatment trials. *Am J Gastroenterol* 1996; 91:277-83.
43. Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol* 2000; 136:381-5.
44. Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials in HEPATOLOGY: predictors of quality. *Hepatology* 1999; 30:1134-8.
45. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005; 365:1159-62.
46. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *Br Med J* 2010; 340:c723.
47. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *Br Med J* 1999; 319:670-4.
48. Lai TY, Wong VW, Lam RF, Cheng AC, Lam DS, Leung GM. Quality of reporting of key methodological items of randomized controlled trials in clinical ophthalmic journals. *Ophthalmic Epidemiol* 2007;14:390-8
49. Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le LJ, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; 347:363-6.
50. Junker CA. Adherence to published standards of reporting: a comparison of placebo-controlled trials published in English or German. *J Am Med Assoc* 1998;280:247-9
51. DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982; 306:1332-7.
52. A proposal for structured reporting of randomized controlled trials. The Standards of Reporting Trials Group. *J Am Med Assoc* 1994; 272:1926-31.

53. Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. *Ann Intern Med* 1994; 121:894-5.
54. Rennie D. Reporting randomized controlled trials. An experiment and a call for responses from readers. *J Am Med Assoc* 1995; 273:1054e5.
55. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *J Am Med Assoc* 1996; 276:637-9.
56. Campbell MK, Piaggio G, Elbourne DR, Altman DG; for the CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *BMJ*. 2012;345:e5661
57. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG, for the CONSORT Group. Reporting of noninferiority and equivalence randomized trials. Extension of the CONSORT 2010 statement. *JAMA*. 2012; 308(24): 2594-2604.
58. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D; CONSORT group; Pragmatic Trials in Healthcare (PractiHC) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*. 2008;337:a2390.
59. Ioannidis JPA, Evans SJW, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D, for the CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT Statement. *Ann Intern Med*. 2004;141(10):781-788.
60. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P; CONSORT NPT Group. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. *Ann Intern Med*. 2017;167(1):40-47.
61. Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C, for the CONSORT Group. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT Statement. *Ann Intern Med*. 2006; 144(5):364-367.
62. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, Schulz KF, the CONSORT Group. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet*. 2008;371(9609):281-283
63. Davidoff F. News from the International Committee of Medical Journal Editors. *Ann Intern Med* 2000;133:229-31.
64. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet* 2001; 357:1191-4.

65. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, for the CONSORT Group. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trial. *BMJ*. 2010;340:c869.
66. Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* 2004; 329:883.
67. Lee W, Bindman J, Ford T, et al. Bias in psychiatric case-control studies: literature survey. *Br J Psychiatry* 2007; 190: 204–09.
68. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. *Am J Epidemiol* 2005; 161: 280–88.
69. Anon. Guidelines for documentation of epidemiologic studies. Epidemiology Work Group of the Interagency Regulatory Liaison Group. *Am J Epidemiol* 1981; 114: 609–13.
70. Rennie D. CONSORT revised – improving the reporting of randomized trials. *JAMA* 2001; 285: 2006–07.
71. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007; 147(8):573-577.
72. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, et al. STrengthening the REporting of Genetic Association Studies. STrengthening the REporting of Genetic Association Studies (STREGA): An Extension of the STROBE Statement. *PLoS Med*. 2009;6(2):e22.
73. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. *Lancet Infect Dis*. 2014. pii: S1473-3099(13)70324-4.
74. Tacconelli E, Cataldo MA, Paul M, Leibovici L, Kluytmans J, Schröder W, et al. STROBE-AMS: recommendations to optimise reporting of epidemiological studies on antimicrobial resistance and informing improvement in antimicrobial stewardship. *BMJ Open*. 2016;6(2):e010134.
75. Equator network : Enhancing the QUALity and Transparency Of health Research Accessible à l'URL: <http://www.equator-network.org/>
76. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *JAMA*. 1994; 272: 1367–1371.

77. Swingler GH, Volmink J, Ioannidis JP. Number of published systematic reviews and global burden of disease: Database analysis. *BMJ*. 2003; 327: 1083–1084.
78. Mulrow CD. The medical review article: State of the science. *Ann Intern Med*. 1987; 106: 485–488.
79. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Metaanalysis of randomized controlled trials. *New Engl J Med*. 1987; 316: 450–455.
80. Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: An update. *Mt Sinai J Med* 1996; 63: 216–224.
81. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, et al. Improving the quality of reporting of meta-analysis of randomized controlled trials: The QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet* 1999; 354: 1896–1900.
82. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *BMJ*. 2009; 339:b2535.
83. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, Moher D, Vohra S; PRISMA harms group. PRISMA harms checklist: improving harms reporting in systematic reviews. *BMJ*. 2016; 352:i157.
84. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, Tierney JF; PRISMA-IPD Development Group. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *JAMA*. 2015;313(16):1657-1665.
85. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM; and the PRISMA-DTA Group. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA*. 2018;319(4):388-396.
86. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*. 2005;93(4):387-391.
87. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D, Burchill SA. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *Br J Cancer*. 2003; 88(8): 1191–1198.
88. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* 2005; 97 (14): 1043–1055.
89. Sigounas DE, Tatsioni A, Christodoulou DK, Tsianos EV, Ioannidis JP. New Prognostic Markers for Outcome of Acute Pancreatitis: Overview of Reporting in 184 Studies. *Pancreas* 2011 40(4): 522–532.

90. Riley DS, Barber MS, Kienle GS, Aronson JK, von Schoen-Angerer T, et al. CARE guidelines for case reports: explanation and elaboration document. *J Clin Epidemiol*. 2017. pii: S0895-4356(17)30037-9.
91. Vandembroucke JP. In Defense of Case Reports and Case Series. *Ann Intern Med* 2001; 134:330e4.
92. Aronson JK. Anecdotes as evidence. *BMJ* 2003;326:1346.
93. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol*. 2012;12(1):181.
94. Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, et al. SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. *Ann Intern Med*. 2013;158(3):200-207
95. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol*. 2010 ;8(6):e1000412.
96. Sharp MK, Tokalić R, Gómez G, Wager E, Altman DG, Hren D. A cross-sectional bibliometric study showed suboptimal journal endorsement rates of STROBE and its extensions. *J Clin Epidemiol*. 2018. pii: S0895-4356(18)30537-7.
97. Jin Y, Sanger N, Shams I, Luo C, Shahid H, Li G, et al. Does the medical literature remainin adequately described despite having reporting guidelines for 21 years? – A systematic review of reviews: an update. *J Multidiscip Healthc*. 2018; 11:495-510.
98. Wang M, Jin Y, Hu ZJ, Thabane A, Dennis B, et al. The reporting quality of abstracts of stepped wedge randomized trials is suboptimal: A systematic survey of the literature. *Contemp Clin Trials Commun*.2017; 8:1-10.
99. Favier R, Crépin S. The reporting of harms in publications on randomized controlled trials funded by the "Programme Hospitalier de Recherche Clinique," a French academic funding scheme. *Clin Trials*. 2018; 15(3):257-267.
100. Candy B, Vickerstaff V, Jones L, King M. Description of complex interventions: analysis of changes in reporting in randomised trials since 2002. *Trials*. 2018; 19(1):110.
101. Chhapola V, Tiwari S, Brar R, Kanwal SK. Reporting quality of trial abstracts-improved yet suboptimal: A systematic review and meta-analysis. *J Evid Based Med*. 2018; 11(2):89-94.

102. Page MJ, Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. *Syst Rev.* 2017;6(1):263.
103. Hutton B, Wolfe D, Moher D, Shamseer L. Reporting guidance considerations from a statistical perspective: overview of tools to enhance the rigour of reporting of randomised trials and systematic reviews. *Evid Based Ment Health.* 2017; 20(2):46-52.
104. Mercieca-Bebber R, Rouette J, Calvert M, King MT, McLeod L, Holch P, et al International Society for Quality of Life Research (ISOQOL) Best Practice for PROs—Reporting Taskforce.. Preliminary evidence on the uptake, use and benefits of the CONSORT-PRO extension. *Qual Life Res.* 2017; 26(6):1427-1437.
105. Nedovic D, Panic N, Pastorino R, Ricciardi W, Boccia S. Evaluation of the Endorsement of the STrengthening the REporting of Genetic Association Studies (STREGA) Statement on the Reporting Quality of Published Genetic Association Studies. *J Epidemiol.* 2016 ;26(8):399-404.
106. Martin J, Taljaard M, Girling A, Hemming K. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. *BMJ Open.* 2016;6(2):e010166.
107. Li J, Gao W, Punja S, Ma B, Vohra S, et al. Reporting quality of N-of-1 trials published between 1985 and 2013: a systematic review. *J Clin Epidemiol.* 2016; 76:57-64.
108. Pouwels KB, Widyakusuma NN, Groenwold RH, Hak E. Quality of reporting of confounding remained suboptimal after the STROBE guideline. *J Clin Epidemiol.* 2016;69:217-24.
109. Hopewell S, Boutron I, Altman DG, Ravaud P. Deficiencies in the publication and reporting of the results of systematic reviews presented at scientific medical conferences. *J Clin Epidemiol.* 2015;68(12):1488-95.
110. Rutterford C, Taljaard M, Dixon S, Copas A, Eldridge S. Reporting and methodological quality of sample size calculations in cluster randomized trials could be improved: a review. *J Clin Epidemiol.* 2015;68(6):716-23.
111. Cui Q, Tian J, Song X, Yang K. Does the CONSORT checklist for abstracts improve the quality of reports of randomized controlled trials on clinical pathways? *J Eval Clin Pract.* 2014;20(6):827-33.

112. Li JL, Ge L, Ma JC, Zeng QL, Yao L, et al. Quality of reporting of systematic reviews published in "evidence-based" Chinese journals. *Syst Rev.* 2014; 3:58.
113. Ge L, Wang JC, Li JL, Liang L, An N, et al. The assessment of the quality of reporting of systematic reviews/meta-analyses in diagnostic tests published by authors in China. *PLoS One.* 2014; 9(1):e85908.
114. Korevaar DA, van Enst WA, Spijker R, Bossuyt PM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med.* 2014;19(2):47-54.
115. Hodkinson A, Kirkham JJ, Tudur-Smith C, Gamble C. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. *BMJ Open.* 2013; 3(9):e003436.
116. Samaan Z, Mbuagbaw L, Kosa D, Borg Debono V, Dillenburg R, et al . A systematic scoping review of adherence to reporting guidelines in health care literature. *J Multidiscip Healthc.* 2013;6:169-88.
117. Clark L, Schmidt U, Tharmanathan P, Adamson J, Hewitt C, et al. Poor reporting quality of key Randomization and Allocation Concealment details is still prevalent among published RCTs in 2011: a review. *J Eval Clin Pract.* 2013;19(4):703-7.
118. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, et al . Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev.* 2012; 11:MR000030.
119. Kim MR, Kim MY, Kim SY, Hwang IH, Yoon YJ. The quality of reporting of cohort, case-control studies in the korean journal of family medicine. *Korean J Fam Med.* 2012; 33(2):79-88.
120. Ghimire S, Kyung E, Kang W, Kim E. Assessment of adherence to the CONSORT statement for quality of reports on randomized controlled trial abstracts from four high-impact general medical journals. *Trials.* 2012; 13:77.
121. Willis BH, Quigley M. The assessment of the quality of reporting of meta-analyses in diagnostic research: a systematic review. *BMC Med Res Methodol.* 2011;11:163.
122. Poorolajal J, Cheraghi Z, Irani AD, Rezaeian S. Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. *Epidemiol Health.* 2011;33:e2011005.

123. Dasí F, Navarro-García MM, Jiménez-Heredia M, Magraner J, Viña JR, et al . Evaluation of the quality of publications on randomized clinical trials using the Consolidated Standards of Reporting Trials (CONSORT) statement guidelines in a Spanish tertiary hospital. *J Clin Pharmacol.* 2012;52(7):1106-14.
124. Haidich AB, Birtsou C, Dardavessis T, Tirodimos I, Arvanitidou M. The quality of safety reporting in trials is still suboptimal: survey of major general medical journals. *J Clin Epidemiol.* 2011;64(2):124-35.
125. Jin Z, Yu D, Zhang L, Meng H, Lu J, et al. A retrospective survey of research design and statistical analyses in selected Chinese medical journals in 1998 and 2008. *PLoS One.* 2010;5(5):e10822.
126. Yu LM, Chan AW, Hopewell S, Deeks JJ, Altman DG. Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review. *Trials.* 2010 ;11:59.
127. Reveiz L, Cortés-Jofré M, Asenjo Lobos C, Nicita G, Ciapponi A, et al ; Iberoamerican Cochrane Network.. Influence of trial registration on reporting quality of randomized trials: study from highest ranked journals. *J Clin Epidemiol.* 2010;63(11):1216-22. Erratum in: *J Clin Epidemiol.* 2011;64(1):116.
128. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ.* 2010 ;340:c723.
129. Uetani K, Nakayama T, Ikai H, Yonemoto N, Moher D. Quality of reports on randomized controlled trials conducted in Japan: evaluation of adherence to the CONSORT statement. *Intern Med.* 2009;48(5):307-13.
130. Arnaud-Coffin P, Maillet D, Gan HK, Stelmes JJ, You B, et al. A systematic review of adverse events in randomized trials assessing immune checkpoint inhibitors. *Int J Cancer.* 2019. [Epub ahead of print]
131. Ter Veer E, van Kleef JJ, Sprangers MAG, Haj Mohammad N, van Oijen MGH, et al . Reporting of health-related quality of life in randomized controlled trials involving palliative systemic therapy for esophagogastric cancer: a systematic review. *Gastric Cancer.* 2018;21(2):183-195.

132. Tardy MP, Gal J, Chamorey E, Almairac F, Vandenbos F, Bondiaet al. Quality of Randomized Controlled Trials Reporting in the Treatment of Adult Patients with High-Grade Gliomas. *Oncologist*. 2018;23(3):337-345.
133. Chen M, Cui J, Zhang AL, Sze DM, Xue CC, et al. Adherence to CONSORT Items in Randomized Controlled Trials of Integrative Medicine for Colorectal Cancer Published in Chinese Journals. *J Altern Complement Med*. 2018; 24(2):115-124.
134. Chen YP, Chen L, Li WF, Lee AWM, Vermorken JB, et al. Reporting Quality of Randomized, Controlled Trials Evaluating Combined Chemoradiotherapy in Nasopharyngeal Carcinoma. *Int J Radiat Oncol Biol Phys*. 2017;98(1):170-176.
135. Ge L, Tian JH, Li XX, Song F, Li L, et al . Epidemiology Characteristics, Methodological Assessment and Reporting of Statistical Analysis of Network Meta-Analyses in the Field of Cancer. *Sci Rep*.2016;6:37208.
136. Aguiar PM, Lima TM, Storpirtis S. Systematic review of the economic evaluations of novel therapeutic agents in multiple myeloma: what is the reporting quality? *J Clin Pharm Ther*. 2016;41(2):189-97.
137. Nayan M, Jayalath VH, Jewett MA, Bedard PL, Hamilton RJ. Randomized controlled trials in testicular cancer: A demographic and quality assessment. *Urol Oncol*. 2016;34(2):60.e7-13.
138. Sivendran S, Newport K, Horst M, Albert A, Galsky MD. Reporting quality of abstracts in phase III clinical trials of systemic therapy in metastatic solid malignancies. *Trials*. 2015 Aug 8;16:341.
139. Altwaairgi AK, Alfakeeh AH, Hopman WM, Parulekar WR. Quality of reporting of chemotherapy compliance in randomized controlled trials of breast cancer treatment. *Jpn J Clin Oncol*. 2015; 45(6):520-6.
140. Huang D, Jin X, Gao J, Li Y, Lu L, et al. Quality evaluation of randomized controlled trials reports of laparoscopy compared with open colorectal resection for colorectal cancer. *Expert Rev Anticancer Ther*. 2015;15(6):727-32.
141. Chen TW, Razak AR, Bedard PL, Siu LL, Hansen AR. A systematic review of immune-related adverse event reporting in clinical trials of immune checkpoint inhibitors. *Ann Oncol*. 2015;26(9):1824-9.
142. Pinzon MC, Hayden DM, Ariel D, Bartosiak KA, Chiodo MV, et al. Are our publications failing the inspection? a review of the publications in rectal cancer surgery between 2002 and 2012. *Dis Colon Rectum*.2014; 57(8):983-92.

143. Carlton DA, Kocherginsky M, Langerman AJ. A systematic review of the quality of randomized controlled trials in head and neck oncology surgery. *Laryngoscope*. 2015; 125(1):146-52.
144. Guo JW, Iribarren SJ. Reporting quality for abstracts of randomized controlled trials in cancer nursing research. *Cancer Nurs*. 2014; 37(6):436-44.
145. Lu L, Liao M, Zeng J, He J. Quality of reporting and its correlates among randomized controlled trials on acupuncture for cancer pain: application of the CONSORT 2010 Statement and STRICTA. *Expert Rev Anticancer Ther*. 2013;13(4):489-98.
146. Vera-Badillo FE, Shapiro R, Ocana A, Amir E, Tannock IF. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. *Ann Oncol*. 2013;24(5):1238-44.
147. Péron J, Pond GR, Gan HK, Chen EX, Almufti R, et al. Quality of reporting of modern randomized controlled trials in medical oncology: a systematic review. *J Natl Cancer Inst*. 2012; 104(13):982-9.
148. Lu L, Zeng J, Chen Y. Quality of reporting in randomized controlled trials conducted in China on the treatment of cancer pain. *Expert Rev Anticancer Ther*. 2011;11(6):871-7.
149. Toulmonde M, Bellera C, Mathoulin-Pelissier S, Debled M, Bui B, Italiano A. Quality of randomized controlled trials reporting in the treatment of sarcomas. *J Clin Oncol*. 2011; 29(9):1204-9.
150. Papatheasiou AA, Zintzaras E. Assessing the quality of reporting of observational studies in cancer. *Ann Epidemiol*. 2010;20(1):67-73.
151. Ziogas DC, Zintzaras E. Analysis of the quality of reporting of randomized controlled trials in acute and chronic myeloid leukemia, and myelodysplastic syndromes as governed by the CONSORT statement. *Ann Epidemiol*. 2009;19(7):494-500.
152. Bassani R, Pereira GKR, Page MJ, Tricco AC, Moher D, et al. Systematic reviews in dentistry: Current status, epidemiological and reporting characteristics. *J Dent*. 2019.[Epub ahead of print]
153. Alamri HM, Alharbi F. Quality Assessment of Randomized Clinical Trials Reporting in Endodontic Journals: An Observational Study from 2012 to 2017. *J Endod*. 2018; 44(8):1246-1250.

154. Lee DW, Shin IS. Critical quality evaluation of network meta-analyses in dental care. *J Dent.* 2018; 75:7-11.
155. Reis A, de Geus JL, Wambier L, Schroeder M, Loguercio AD. Compliance of Randomized Clinical Trials in Noncarious Cervical Lesions With the CONSORT Statement: A Systematic Review of Methodology. *Oper Dent.* 2018;43(3):E129-E151.
156. Lucena C, Souza EM, Voinea GC, Pulgar R, Valderrama MJ, et al. A quality assessment of randomized controlled trial reports in endodontics. *Int Endod J.* 2017;50(3):237-250.
157. Kloukos D, Papageorgiou SN, Doulis I, Petridis H, Pandis N. Reporting quality of randomised controlled trials published in prosthodontic and implantology journals. *J Oral Rehabil.* 2015;42(12):914-25.
158. Lempesi E, Koletsi D, Fleming PS, Pandis N. The reporting quality of randomized controlled trials in orthodontics. *J Evid Based Dent Pract.* 2014;14(2):46-52.
159. Kiriakou J, Pandis N, Fleming PS, Madianos P, Polychronopoulou A. Reporting quality of systematic review abstracts in leading oral implantology journals. *J. Dent.* 2013; 41(12):1181-7.
160. Savithra P, Nagesh LS. Have CONSORT guidelines improved the quality of reporting of randomised controlled trials published in public health dentistry journals? *Oral Health Prev Dent.* 2013; 11(2):95-103.
161. Fleming PS, Buckley N, Seehra J, Polychronopoulou A, Pandis N. Reporting quality of abstracts of randomized controlled trials published in leading orthodontic journals from 2006 to 2011. *Am J Orthod Dentofacial Orthop.* 2012;142(4):451-8.
162. Xiang C, Li C, Dong W, Zhang L, Shi Z, et al. Assessment of reporting quality of randomized controlled trial related to dentine hypersensitivity. *Hua Xi Kou Qiang Yi Xue Za Zhi.* 2012;30(3):267-70, 274. Chinese.
163. Meijer HJ, Raghoobar GM. Quality of reporting of descriptive studies in implant dentistry. Critical aspects in design, outcome assessment and clinical relevance. *J Clin Periodontol.* 2012;39 Suppl 12:108-13.
164. Pjetursson BE, Zwahlen M, Lang NP. Quality of reporting of clinical studies to assess and compare performance of implant-supported restorations. *J Clin*

Periodontol. 2012 ;39 Suppl 12:139-59.

165. Froud R, Eldridge S, Diaz Ordaz K, Marinho VC, Donner A. Quality of cluster randomized controlled trials in oral health: a systematic review of reports published between 2005 and 2009. *Community Dent Oral Epidemiol.* 2012;40 Suppl 1:3-14.
166. Richards D. Quality of reporting randomised controlled trials in dental public health. *Evid Based Dent.* 2011;12(2):54.
167. Marshman Z, Farid F. The quality of reporting of randomised controlled trials in dental public health. *Community Dent Health.* 2010;27(4):253-6.
168. Pandis N, Polychronopoulou A, Eliades T. An assessment of quality characteristics of randomised control trials published in dental journals. *J Dent.* 2010;38(9):713-21.
169. Jia P, Tang L, Yu J, Liu J, Kang D, Sun X. The quality of reporting in randomized controlled trials of acupuncture for knee osteoarthritis: A cross-sectional survey. *PLoS One.* 2018; 13(4):e0195652.
170. You YN, Cho MR, Kim JH, Park JH, Park GC, et al. Assessing the quality of reports about randomized controlled trials of scalp acupuncture combined with another treatment for stroke. *BMC Complement Altern Med.* 2017; 17(1):452.
171. Zeng J, Lin G, Li L, Lu L, Chen C, et al. Assessment of reporting quality in randomised controlled trials of acupuncture for post-stroke rehabilitation using the CONSORT statement and STRICTA guidelines. *Acupunct Med.* 2017 ;35(2):100-106.
172. Wen W, Yang L, Liu S, Zhong Y, Hu X, et al. Reporting quality and effect size comparison in randomized controlled trials of bo's abdominal acupuncture using CONSORT statement and STRICTA. *J Tradit Chin Med.* 2016 ;36(3):382-91. Chinese.
173. Lu LM, He J, Zeng JC, Liao MX, Jia C, et al. Impact evaluation of CONSORT and STRICTA guidelines on reporting quality for randomized controlled trials of acupuncture conducted in China. *Chin J Integr Med.* 2017; 23(1):10-17.
174. Ma B, Chen ZM, Xu JK, Wang YN, Chen KY, et al. Do the CONSORT and STRICTA Checklists Improve the Reporting Quality of Acupuncture and Moxibustion Randomized Controlled Trials Published in Chinese Journals? A Systematic Review and Analysis of Trends. *PLoS One.* 2016;11(1):e0147244.

175. Yu H, Wang Y, Guo Y, Wang H, Chen B, et al. Quality assessment of randomized controlled trials reporting on knee osteoarthritis treated with warming needle moxibustion. *J Tradit Chin Med*. 2014; 34(5):621-6.
176. Liu Y, Zhang R, Huang J, Zhao X, Liu D, et al. Reporting quality of systematic reviews/meta-analyses of acupuncture. *PLoS One*. 2014;9(11):e113172.
177. Kim KH, Kang JW, Lee MS, Lee JD. Assessment of the quality of reporting in randomised controlled trials of acupuncture in the Korean literature using the CONSORT statement and STRICTA guidelines. *BMJ Open*. 2014;4(7):e005068.
178. Zhuang L, He J, Zhuang X, Lu L. Quality of reporting on randomized controlled trials of acupuncture for stroke rehabilitation. *BMC Complement Altern Med*. 2014;14:151.
179. Bo C, Xue Z, Yi G, Zelin C, Yang B, et al. Assessing the quality of reports about randomized controlled trials of acupuncture treatment on Diabetic Peripheral Neuropathy. *PLoS One*. 2012;7(7):e38461.
180. Colbert AP, Spaulding K, Larsen A, Ahn AC, Cutro JA. Electrodermal activity at acupoints: literature review and recommendations for reporting clinical trials. *J Acupunct Meridian Stud*. 2011;4(1):5-13.
181. Lu X, Hongcai S, Jiaying W, Jing H, Jun X. Assessing the quality of reports about randomized controlled trials of acupuncture treatment on mild cognitive impairment. *PLoS One*. 2011;6(2): e16922.
182. Zheng S, Du J, Lu X, Zhang Y, Hu L, et al. Quality of randomized controlled trials in acupuncture treatment of hepatitis B virus infection--a systematic review. *Acupunct Electrother Res*. 2010;35(3-4):119-31.
183. Hendarto A, Yosia M, Sastroasmoro S. Quality analysis of randomized controlled trials reporting in the field of pediatrics by Indonesian researchers. *J Evid Based Med*. 2018 [Epub ahead of print]
184. Devos F, Ibrahim N, Foissac F, Bouazza N, Ancel PY, et al. Comparison of the Quality of Pediatric Randomized Controlled Trials Published in Both Nursing and Medical Journals: Adherence to the CONSORT Statement. *Worldviews Evid Based Nurs*. 2018;15(6):447-454.
185. Pansa P, Hsia Y, Bielicki J, Lutsar I, Walker AS, et al. Evaluating Safety Reporting in Paediatric Antibiotic Trials, 2000-2016: A Systematic Review and Meta-Analysis. *Drugs*. 2018;78(2):231-244.

186. Braga LH, McGrath M, Easterbrook B, Jegatheeswaran K, Mauro L, et al. Quality of reporting for randomized controlled trials in the hypospadias literature: Where do we stand? *J Pediatr Urol*. 2017; 13(5):482.e1-482.e9.
187. Braga LH, Lorenzo AJ, Bagli DJ, Pippi Salle JL, Caldamone A. Application of the STROBE statement to the hypospadias literature: Report of the international pediatric urology task force on hypospadias. *J Pediatr Urol*. 2016; 12(6):367-380.
188. Chhapola V, Tiwari S, Brar R, Kanwal SK. An interrupted time series analysis showed suboptimal improvement in reporting quality of trial abstract. *J Clin Epidemiol*. 2016; 71:11-7.
189. McGee RG, Craig JC, Rogerson TE, Webster AC. Systematic reviews of surgical procedures in children: quantity, coverage and quality. *J Paediatr Child Health*. 2013;49(4):319-24.
190. Ma B, Ke FY, Chen ZM, Qi GQ, Li HM, et al. Does the reporting of randomized clinical trials published in Chinese pediatrics journals improve after the CONSORT Statement is adopted? *Contemp Clin Trials*. 2012;33(5):889-94.
191. Brooks RJ, Higgins GY, Webster AC. Systematic review of randomized controlled trial quality in pediatric kidney transplantation. *Pediatr Nephrol*. 2010;25(12):2383-92.
192. Al Faleh K, Al-Omran M. Reporting and methodologic quality of Cochrane Neonatal review group systematic reviews. *BMC Pediatr*. 2009; 9:38.
193. Grellety T, Petit-Moneger A, Diallo A, Mathoulin-Pelissier S, Italiano A. Quality of reporting of phase II trials: a focus on highly ranked oncology journals. *Ann Oncol* 2014; 25:536-41.
194. Ottaiano A, Mollo E, Di Lorenzo G, Pisano C, Di Maio M, Barletta E, et al. Prospective clinical trials of biotherapies in solid tumors: a 5-year survey. *Cancer Immunother*. 2005;54(1):44-50.
195. Thezenas S, Duffour J, Culine S, Kramar A. Five-year change in statistical designs of phase II trials published in leading cancer journals. *Eur J Cancer*. 2004;40(8):1244-9
196. Nickolich M, Babakoochi S, Fu P, Dowlati A. Clinical trial design in small cell lung cancer: surrogate end points and statistical evolution. *Clin Lung Cancer*.

2014;15(3):207–12.

197. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ*. 2004 ;328(7430) :22–4.
198. Perrone F, Di Maio M, De Maio E, Maione P, Ottaiano A, Pensabene M, et al. Statistical design in phase II clinical trials and its application in breast cancer. *Lancet Oncol*. 2003;4(5):305–11.
199. Rivoirard R, Bourmaud A, Oriol M, Tinquaut F, Méry B, Langrand-Escure J, et al. Quality of reporting in oncology studies: A systematic analysis of literature reviews and prospects. *Crit Rev Oncol Hematol*. 2017 Apr; 112:179-189.
200. Langrand-Escure J, Rivoirard R, Oriol M, Tinquaut F, Rancoule C, Chauvin F, et al. Quality of reporting in oncology phase II trials: A 5-year assessment through systematic review. *PLoS One*. 2017 Dec 7;12(12): e0185536.
201. Rivoirard R, Langrand-Escure J, Oriol M, Tinquaut F, Chauvin F, Rancoule C, et al. Evaluation of the quality of the reporting of phase II clinical trials in oncology: A systematic review. *Crit Rev Oncol Hematol*. 2018 May;1 25:78-83.
202. Walter, T., Krzyzanowska, M.K. Quality of clinical trials in Gastroenteropancreatic neuroendocrine tumours. *Neuroendocrinology* 2012 ;96 :238–248
203. El-Maraghi, R.H., Eisenhauer, E.A. Review of phase II trial designs used in studies of molecular targeted agents: outcomes and predictors of success in phase III. *J. Clin. Oncol*. 2008 ;26 (8) :1346–1354.
204. Booto Ekionea JP, Bernard P, Plaisent M. Consensus par la méthode Delphi sur les concepts clés des capacités organisationnelles spécifiques de la gestion des connaissances. : Accessible à l'URL : [http://www.recherche-qualitative.qc.ca/documents/files/revue/edition_reguliere/numero29\(3\)/RQ_29\(3\)_Booto-et-al.pdf](http://www.recherche-qualitative.qc.ca/documents/files/revue/edition_reguliere/numero29(3)/RQ_29(3)_Booto-et-al.pdf)
205. NIH : U.S. National Library of Medicine ; Clinicaltrials.gov : Accessible à l'URL : <https://clinicaltrials.gov/>
206. Ivanova A, Paul B, Marchenko O, Song G, Patel N, Moschos SJ. Nine-year change in statistical design, profile, and success rates of Phase II oncology trials.

- J Biopharm Stat. 2016 ;26 :141-9.
207. Korn EL, Freidlin B. Interim futility monitoring assessing immune therapies with a potentially delayed treatment effect. *J Clin Oncol* 2018 ;36 :2444–2449
 208. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple disease, or both. *N Engl J Med.*2017 ; 377 :62–70.
 209. Yuan Y, Guo B, Munsell M, et al. MIDAS : a practical Bayesian design for platform trials with molecularly targeted agents. *Stat Med.* 2016 ; 35 : 3892–3906.
 210. Lin R, Coleman RL, Yuan Y. TOP : Time-to-Event Bayesian Optimal Phase II Trial Design for Cancer Immunotherapy. *J Natl Cancer Inst.* 2019 ; 29. pii: djz049
 211. Zhou H, LeeJJ, YuanY. BOP2 : Bayesian optimal design for phase II clinical trials with simple and complex endpoints. *Stat Med.* 2017 ; 36 :3302–3314.
 212. Thall PF, Simon R. Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics.*1994 ;5 :337–349.
 213. Nicole Gray. Success rate for clinical trials climbed in recent years: study. Accessible à l'URL : <https://www.biopharmadive.com/news/success-rate-for-clinical-trials-climbed-in-recent-years-study/420973/>
 214. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004; 3:711-716.
 215. DiMasi JA, Grabowski HG. Economics of New Oncology Drug Development. *J Clin Oncol* 2007 ;25 :209-16.
 216. Rivoirard R, Duplay V, Oriol M, Tinquaut F, Chauvin F, Magne N et al. Outcomes Definitions and Statistical Tests in Oncology Studies: A Systematic Review of the Reporting Consistency. *PLoS One.* 2016 Oct 7;11(10): e0164275.

