



Évaluation d'outils d'accès à la connaissance en médecine générale

Matthieu Schuers

► To cite this version:

Matthieu Schuers. Évaluation d'outils d'accès à la connaissance en médecine générale. Médecine humaine et pathologie. Normandie Université, 2017. Français. NNT : 2017NORMR070 . tel-01887674

HAL Id: tel-01887674

<https://theses.hal.science/tel-01887674>

Submitted on 4 Oct 2018

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

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



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité Sciences de la Vie et de la Santé

Préparée au sein de l'Université de Rouen Normandie

Evaluation d'outils d'accès à la connaissance en médecine générale

**Présentée et soutenue par
Matthieu SCHUERS**

**Thèse soutenue publiquement le 19 octobre 2017
devant le jury composé de**

M. Stéfan DARMONI	PU-PH, Université de Rouen Normandie	Directeur de thèse
M. Pascal STACCINI	PU-PH, Université de Nice Sophia Antipolis	Rapporteur
M. Olivier SAINT LARY	PU-MG, Université Versailles Saint Quentin	Rapporteur
Mme Isabelle AUBIN-AUGER	PU-MG, Université Paris VII	Examinatrice
M. Nicolas GRIFFON	PH, CHU de Rouen	Examineur
M. Alain MERCIER	PU-MG, Université Paris XIII	Examineur
M. Olivier STEICHEN	PU-PH, Université Paris VI	Examineur

**Thèse dirigée par Stéfan DARMONI,
Laboratoire LITIS (Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes)**



« Tu comprends, sur cette terre, y a une chose effroyable : c'est que tout le monde a ses raisons... ».

La Règle du Jeu
Jean RENOIR, 1939

Remerciements

Au Pr Stéfan DARMONI,

Merci pour m'avoir accueilli dans votre équipe, pour m'avoir accordé votre confiance (et bien plus) sans même me connaître et pour avoir dirigé cette thèse. Jour après jour, et parfois contre vents et marées, vous insufflez une quantité phénoménale de vie aux membres de votre équipe (consentants ou non). Dans un milieu normé, vous possédez l'immense mérite d'être différent.

Au Pr Pascal STACCINI,

Merci d'avoir accepté d'être mon rapporteur et d'avoir évalué ce travail.

Au Pr Olivier SAINT LARY,

Merci d'avoir accepté d'être mon rapporteur et d'avoir évalué ce travail. Merci surtout pour ton implication quotidienne et sans relâche au service de la discipline. Je sais les sacrifices qu'elle implique, et t'en suis profondément reconnaissant.

Au Pr Isabelle AUBIN-AUGER,

Merci d'avoir accepté de faire partie du jury de cette thèse et d'avoir évalué ce travail. Tu as fait partie, avec Alain, des pionniers de la recherche en médecine générale. J'espère que tu mesures l'ampleur des vocations que vous avez contribués à créer. Merci pour ta rigueur et ta bienveillance sans faille.

Au Pr Alain MERCIER,

Merci d'avoir accepté de faire partie du jury de cette thèse et d'avoir évalué ce travail. Merci de l'avoir initié, surtout, et d'y avoir largement participé. Merci de m'avoir, il y a bientôt dix ans, transmis le virus de la recherche. J'espère avoir été à la hauteur.

Au Pr Olivier STEICHEN,

Merci d'avoir accepté de faire partie du jury de cette thèse et d'avoir évalué ce travail.

Au Dr Nicolas GRIFFON,

Merci de m'avoir accompagné tout au long de ces quatre années passées ensemble dans l'équipe du D2IM. Te connaissant, tu sous-estimes probablement l'étendue de ta contribution. Je reste admiratif de ta rigueur et de ton éthique scientifiques. Je te remercie de ta présence au sein de ce jury et espère avoir encore de nombreuses occasions de travailler avec toi (à condition qu'il ne soit plus jamais question de délai total ou de *turnaround time*).

A tous les membres du D2IM,

Merci de m'avoir accueilli au sein de votre équipe, qui réussit le tour de force d'être à la fois diverse et unie. Merci d'avoir partagé votre enthousiasme, votre humour, votre expertise, et votre plaisir à travailler ensemble.

Au Pr Jean-Loup HERMIL,

Merci pour ton implication quotidienne au service de la médecine générale universitaire et pour ta bienveillance envers les étudiants. Merci de m'accorder ta confiance, j'espère en être digne.

A tous les membres du DUMG,

Merci de m'avoir soutenu tout au long de ce projet, de m'avoir permis de m'y consacrer pleinement. Sa concrétisation vous doit beaucoup.

Au Dr Jacques FRICHET,

Merci d'avoir proposé de m'accueillir, il y a 9 ans, à ma sortie de la faculté. Merci de garder une foi inébranlable dans la force de ton projet et du collectif qui le fait exister. Tu as su concrétiser une organisation que beaucoup ne parvenaient même pas à imaginer. Comme le dit le célèbre adage, « impossible n'est pas Frichet ».

A toute l'équipe de la maison de santé,

Merci de m'avoir accueilli parmi vous, il y a bientôt dix ans. Votre implication quotidienne au service de la santé de la population, votre engagement collectif sont admirables. Je suis fier de travailler à vos côtés.

A ma famille et mes amis, et tous ceux qui sont un peu des deux. Merci de me supporter, certains depuis bien longtemps. Je ne vous promets pas de grands changements pour l'avenir, mais, une chose est sûre, j'arrête de vous casser les pieds avec cette thèse.

A Kty, puisqu'il est d'usage de finir par le(la) meilleur(e). Depuis janvier 2005, nos chemins syndicaux, associatifs, académiques et même privés ont fait plus que se croiser : ils tracent un sillon étonnamment parallèle. Depuis 13 ans, nous continuons d'avancer, à des vitesses différentes, mais sans jamais se perdre de vue. Je ne sais pas trop comment ça s'appelle (tu me connais), mais je sais que c'est important. Parce que rien n'aurait été possible sans toi, à toi, la reine du link, de l'apéro, des discussions sans fin, des remises en question, des projets fous (« Mais Kty, ça marchera JAMAIS ! »), bref notre reine à tous, je dédie cette thèse.

Les médecins, et en particulier les médecins généralistes, sont confrontés quotidiennement à la difficulté de ne pas tout savoir. En médecine générale, l'étendue du champ d'activité impose aux professionnels une maîtrise de plus en plus aiguë de l'information, qui doit s'appuyer sur le développement d'une littératie numérique, c'est-à-dire la capacité d'utiliser et de comprendre les outils et médias numériques. Les objectifs de ce travail étaient d'identifier les obstacles à la recherche d'information chez les médecins généralistes et de développer et d'évaluer des outils susceptibles de surmonter ces difficultés.

Pour cela, nous avons réalisé une étude qualitative auprès d'internes de médecine générale et de médecins généralistes français. Elle a permis de mettre en évidence de nombreux obstacles à la recherche d'informations en santé en ligne. Ces obstacles concernent les professionnels et les outils, certains sont également d'ordre institutionnel. Parmi ces obstacles figure la langue anglaise, qui semble constituer un repoussoir pour les médecins installés mais également pour les médecins en formation. D'après les données que nous avons recueillies, ce frein concerne aussi bien l'information en elle-même que son support.

Plusieurs des outils développés par l'équipe du Département d'Informatique et d'Information Médicales du CHU de Rouen ont vocation à répondre à ces difficultés. La base de données bibliographiques LiSSa permet l'accès à plusieurs centaines de milliers de références francophones en santé, issues de PubMed mais également de revues francophones non indexées dans PubMed. Le constructeur de requêtes bibliographiques médicales permet à des utilisateurs non anglophones de construire des équations de recherche complexes dans leur langue maternelle, leur permettant de requêter dans PubMed, LiSSa ou CISMef.

Mots clés : Comportement de recherche d'information, Médecine générale, Internet.

Physicians, and especially general practitioners, are not « all-knowing ». As they are confronted with a large amount of clinical situations, information and knowledge management are essential topics of interest. This management is mainly based on the development of numeric literacy, that is the ability to use and understand digital tools and media. The aims of this work were to identify the obstacles encountered by general practitioners when seeking health information online and to develop and assess tools to help them to overcome these barriers.

We performed a qualitative study among general practice residents and general practitioners. This allowed us to highlight several obstacles encountered by them when seeking medical information online. These obstacles include both tools and professionals' issues. Institutional obstacles were also identified. Of these barriers, the insufficient command of English was cited as an important issue by both residents and physicians.

The Department of Medical Information and Informatics of the Rouen University Hospital has developed several tools in order to facilitate access to knowledge for health professionals. LiSSa is a bibliographic database containing only articles written in French. It allows health professionals and researchers, whose native language is not English, to access to hundreds of thousands of references, from PubMed and from journals not indexed in PubMed. A multi-lingual query builder to facilitate information retrieval was designed. It appears to be an effective tool to improve the quality of PubMed queries in particular for users whose first native language is not English.

Key words: Information seeking behaviour, General Practice, Internet.

Préambule	8
------------------	----------

1. Introduction générale	12
---------------------------------	-----------

- 1.1. Donnée, information, recherche d'information, besoin d'information
 - 1.1.1. Définitions
 - 1.1.2. La recherche d'information
 - 1.1.2.1. Définitions francophones
 - 1.1.2.2. Définitions anglophones
 - 1.1.3. Le besoin d'information
- 1.2. Médecine générale et recherche d'information en santé en ligne
 - 1.2.1. La médecine fondée sur les preuves
 - 1.2.2. Les besoins en information des médecins généralistes
 - 1.2.2.1. Analyse quantitative des informations utilisées et des questions posées
 - 1.2.2.2. Analyse qualitative des questions posées
 - 1.2.3. Les supports d'information en santé disponibles en ligne pour les professionnels de santé

2. Les comportements des médecins généralistes à la recherche d'informations en santé en ligne	26
---	-----------

- 2.1. De la question à la recherche d'une réponse
- 2.2. Les supports d'information en ligne des médecins généralistes
- 2.3. Les obstacles à la recherche d'information
- 2.4. Attitudes et comportements des internes et médecins généralistes à la recherche d'information sur Internet
- 2.5. Conclusion

3. LiSSA : Littérature Scientifique en Santé	39
---	-----------

- 3.1. Introduction : les enjeux d'une base de données bibliographiques francophone dans le domaine de la santé
 - 3.1.1. La prédominance de la langue anglaise
 - 3.1.2. La disparition progressive de la production francophone
 - 3.1.3. Un niveau de maîtrise de la langue anglaise qui reste insuffisant
 - 3.1.4. Des outils existants qui ne répondent pas aux besoins des utilisateurs francophones
- 3.2. Présentation de la base de données LiSSa
 - 3.2.1. Les données sources
 - 3.2.2. Le modèle de données et le moteur de recherche
 - 3.2.3. Les fonctionnalités de LiSSa

3.2.4. Formation des utilisateurs de LiSSa	
3.2.5. Utilisation et évaluation de l'outil	
3.3. Publications	
3.4. Un moteur de recherche permettant d'accéder aux sous-ensembles monolingues de PubMed. Preuve de concept et évaluation en français.	
4. Le CRBM : Constructeur de Requêtes Bibliographiques Médicales	68
4.1. Requêter dans PubMed dans sa langue maternelle : les enjeux de la création d'un outil francophone	
4.2. Le constructeur de requêtes bibliographiques médicales : description de l'outil et de ses fonctionnalités	
4.3. Données d'utilisation du constructeur de requêtes bibliographiques médicales	
4.4. Publication	
4.5. Un constructeur de requêtes multilingue améliore la qualité des requêtes effectuées dans PubMed : un essai contrôlé randomisé.	
4.6. Conclusion	
5. Thèmes et tendances dans la production scientifique en médecine générale	81
5.1. Introduction	
5.1.1. La recherche en médecine générale	
5.1.2. La bibliométrie	
5.1.3. Analyse bibliométrique de la production scientifique en médecine générale	
5.2. Tendances et thématiques en médecine générale dans PubMed	
5.3. Trends and topics in general practice in PubMed	
6. Conclusion générale	112
7. Bibliographie	115
8. Productions associées à la thèse	124

Préambule

La compétence des professionnels de santé est déterminante pour la qualité et la sécurité des soins délivrés aux patients. Plusieurs institutions et auteurs ont proposé de définir ce qu'est la compétence. Les définitions diffèrent mais se rejoignent sur le fond. Le parlement européen a défini en 1996 la compétence comme l'ensemble des caractéristiques individuelles (connaissances, aptitudes, attitudes) qui permettent à une personne d'exercer son activité de manière autonome, de perfectionner sans cesse sa pratique et de s'adapter à un environnement en mutation rapide (1). Dans une revue de la littérature publiée en 2002, Epstein et Hundert définissent la compétence professionnelle en santé comme l'utilisation habituelle et judicieuse de communication, connaissances, savoir-faire techniques, raisonnement clinique, émotions, valeurs et réflexion dans la pratique de tous les jours pour le bénéfice des individus et de la société (2). Pour Jacques Tardif, il s'agit d'un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes (connaissances acquises, expérience, savoir-faire) et externes (individus, revues médicales, outils en ligne...) à l'intérieur d'une famille de situations (3). Quelle que soit la définition, il apparaît que la compétence en santé est une notion dynamique, s'appuyant notamment sur la capacité des professionnels de santé à gérer l'information.

C'est autour de cette même préoccupation, la gestion de l'information en santé, que s'est développé le Département d'Informatique et d'information Médicales (D2IM) du Centre Hospitalier Universitaire (CHU) de Rouen. Dès 1994, le CHU de Rouen devenait le premier hôpital universitaire français à être raccordé massivement à l'Internet. En février 1995, naissait CISMef : le Catalogue et Index des Sites Médicaux de langue Française, qui décrit et indexe les principales ressources de santé disponibles librement sur Internet en français. Dès 2000, l'équipe pluridisciplinaire du département a développé la première version de son moteur de recherche sémantique (4). Ses domaines de recherche sont l'indexation automatique et la recherche d'information dans un monde mono puis multi-terminologique, puis dans une dimension multilingue. Le département a progressivement développé son propre serveur terminologique, multi-terminologique et inter-lingue

(permettant la navigation matricielle à la fois entre les terminologies et entre les langues), le serveur HeTOP (*Health Terminology-Ontology Portal*, Portail termino-ontologique de santé), devenu depuis 2007 l'outil « fondation » sur lequel repose tous les autres outils du département.

J'ai intégré l'équipe du Département d'Informatique et d'Information Médicales en novembre 2013. Mes objectifs au sein de cette équipe étaient les suivants :

- Développer la « médicalisation » du département, grâce à l'accueil d'internes de médecine générale dans des stages orientés vers la recherche. De novembre 2014 à octobre 2017, sept internes de médecine générale ont été accueillis à mi-temps dans le département.
- Collaborer avec les membres de l'équipe à la mise en place et au suivi de plusieurs projets de recherche en lien avec la thématique des systèmes d'information en santé. Les deux principaux projets concernés étaient le projet EVALSI, qui consistait à évaluer l'impact quantitatif et qualitatif du déploiement des circuits de la biologie et de l'imagerie sur les délais et la qualité des soins et le projet MATRIX, relatif à l'impact du dossier pharmaceutique sur la pratique médicale ainsi que les conditions organisationnelles de réussite d'une généralisation du dossier pharmaceutique auprès des médecins.
- Travailler, dans le cadre d'une thèse de doctorat débutée en novembre 2013, sur la manière dont les médecins généralistes accèdent à la connaissance en santé, et évaluer plusieurs outils développés par le D2IM afin de faciliter l'accès à la connaissance pour ces mêmes professionnels. C'est cette mission qui fait l'objet de ce rapport.

Au cours de ces quatre années de doctorat, la problématique de l'accès à la connaissance des médecins généralistes a été explorée de différentes manières.

Dans un premier temps et dans le cadre d'une étude exploratoire, nous nous sommes intéressés à la manière dont les médecins généralistes, en activité ou en formation initiale, cherchaient de l'information en santé en ligne. Cette étape initiale nous a paru importante, car il nous semblait que les données existantes de la littérature sur les obstacles à la recherche d'information chez les médecins généralistes ne permettaient pas d'expliquer l'ensemble des difficultés rencontrées par les professionnels. Afin de répondre à cette

question, nous avons mené une étude qualitative par focus groups, auprès de médecins généralistes et d'internes de médecine générale de Normandie et d'Ile de France. Les données de cette étude nous ont permis d'identifier les thèmes de recherche des médecins généralistes, leurs sources et supports d'information, leurs modalités de recherche d'information en ligne, les difficultés rencontrées et leurs solutions pour y remédier.

Parmi les obstacles à la recherche d'information identifiés dans l'étude qualitative figurait le fait que de nombreuses sources et supports d'information sont en langue anglaise. Le Département d'Informatique et d'Information Médicales défend depuis sa création le maintien de sources et de supports d'information en langue française, afin de garantir aux professionnels et aux patients non-anglophones l'accès à une information de qualité. La seconde étape de ce travail a concerné deux outils développés par le D2IM :

- LiSSa (Littérature Scientifique en Santé), base de données bibliographiques en santé francophone, regroupant à la fois des références issues de la base de données PubMed et des références issues d'éditeurs comme Elsevier, non indexées dans la base de données états-unienne. Dans ce rapport sont présentés plusieurs étapes du travail du D2IM autour de cet outil. La première étape a consisté à évaluer la qualité en termes de précision et le rappel de la version française d'un moteur de recherche permettant l'accès au sous-ensemble francophone de PubMed. La seconde étape a consisté, via la publication de deux articles francophones, à faire connaître cet outil aux professionnels de santé, essentiellement médicaux.
- Le constructeur de requêtes bibliographiques médicales (CRBM). Il s'agit du premier outil disponible permettant d'élaborer des requêtes complexes en français dans la base de données PubMed. Notre travail a consisté à évaluer l'impact de l'utilisation du CRBM sur la qualité des requêtes effectuées dans PubMed par des internes de médecine générale dont le français était la langue maternelle.

Enfin, dans un dernier temps, nous nous sommes intéressés à la production scientifique en médecine générale. Il nous a semblé possible que les difficultés d'accès à la connaissance des médecins généralistes soient liées au moins en partie à la difficulté d'identifier précisément les informations utiles à leur activité clinique. Quels sont les contours de la production scientifique en médecine générale ? L'objectif de notre travail était de fournir une analyse de l'ensemble de la production scientifique indexée sur

PubMed dans le champ de la médecine générale, afin d'identifier les principaux sujets ainsi que les acteurs prédominants (revues, pays) impliqués dans ce domaine.

Ce rapport présente les différentes parties de ce travail. Il s'articule autour des publications relatives à chacune des thématiques évoquées (cinq publications acceptées dans des revues indexées et un article en cours de relecture).

1. Introduction générale

La capacité d'un professionnel de santé à mobiliser et articuler ses connaissances d'une part et les informations permettant de prendre en charge de manière optimale ses patients d'autre part constitue l'une des bases de la compétence professionnelle (3).

Dans ce chapitre, après avoir défini les notions de donnée, d'information, de connaissance et de recherche d'information, nous tenterons d'identifier les besoins d'information des médecins généralistes et les supports d'information disponibles en ligne pour répondre à ces besoins.

1.1. Donnée, information, recherche d'information, besoin d'information

1.1.1. Définitions

Une **donnée** est une description élémentaire d'une réalité. C'est par exemple une observation ou une mesure. Elle est dépourvue de tout raisonnement, supposition, constatation ou probabilité (5). Il s'agit d'un élément brut qui n'a pas encore été interprété et mis en contexte. C'est là toute la différence entre une donnée et une information : une **information** est une donnée interprétée. En d'autres termes, la mise en contexte d'une donnée crée de la valeur ajoutée pour constituer une information. L'information est un ensemble de données intelligible, qui prend un sens (6).^a

On peut considérer la **connaissance** comme une information comprise, c'est-à-dire assimilée et utilisée, qui permet d'aboutir à une action. L'information en soi n'a donc qu'un intérêt relatif, elle ne vaudra que parce qu'elle sert de marchepied pour accéder à la connaissance (6).

^a Lors d'un recueil de température chez un patient, les mesures « 38,7 », « 39 » et « 39,6 » constituent un ensemble de données. Une fois interprétées et contextualisées, elles fournissent plusieurs informations : le patient est fébrile d'une part, sa température augmente d'autre part.

1.1.2. La recherche d'information

1.1.2.1. Définitions francophones

Les premières définitions francophones de la recherche d'information sont antérieures à l'avènement d'Internet. L'Association Française de Normalisation (AFNOR) distinguait en 1987 la « recherche documentaire », définie comme « Action, méthodes et procédures ayant pour objet de retrouver dans des fonds documentaires les références des documents pertinentes » de la « recherche d'information », définie comme « Action, méthodes et procédures ayant pour objet d'extraire d'un ensemble de documents les informations voulues » (7). Ces définitions correspondent aux modalités traditionnelles de recherche, interrogations distinctes d'un système bibliographique (la recherche documentaire) et d'un système de recherche d'information, banque de données de documents en texte intégral (la recherche d'information). Elles peuvent désormais apparaître comme inappropriées, compte tenu de l'évolution des systèmes d'information, regroupant à l'instar des principaux moteurs du Web les deux types de recherche. De plus, elles méconnaissent le processus de recherche tel que conduit par le chercheur d'information, et il reste difficile d'identifier ce qui appartient à une description des systèmes ou à celle des pratiques. Même si ces définitions ont été actualisées en 2004, elles restent dans les faits peu différenciables (8).

1.1.2.2. Définitions anglophones

La discipline *Library and Information Science* (LIS) se retrouve dans les pays anglo-saxons, les pays nordiques et, de plus en plus, asiatiques. Ses objets de recherche correspondent aux sciences de l'information françaises, mais aussi, et de manière bien plus développée, aux travaux sur l'activité de recherche d'information (8).

La langue anglaise offre plusieurs verbes pour définir l'action de chercher par des humaines : *Information Retrieval*, *Information Searching* et *Information Seeking*. Chaque verbe apporte des nuances permettant de produire des définitions témoignant de l'élargissement de la compréhension de la recherche d'information. L'expression *Information Retrieval* signifie retrouver l'information, au sens de la localiser et de la récupérer. Elle désigne les travaux sur les modélisations des systèmes de recherche d'information. Elle est utilisée dans le terme MeSH *Information Storage and Retrieval*, soit mémorisation et recherche d'information. Les deux expressions *Information Searching* et *Information Seeking* peuvent être traduites par « recherche d'information » ou « recherche

documentaire ». En langue anglaise, la distinction entre les deux verbes n'est pas évidente, mais il semble que *seek* reflète une idée de processus, tentative pour découvrir quelque chose, chercher à atteindre. C'est ainsi que les chercheurs tendent à caractériser à l'aide de l'expression *Information Seeking* le processus entier de recherche d'information, et à utiliser *Information Searching* pour désigner les seules interactions avec les systèmes de recherche d'information (comme formuler des requêtes, par exemple) (8). Ces définitions ont connu encore récemment des évolutions au sein de la discipline LIS, qu'il ne paraît pas utile d'évoquer ici.

Les études issues de l'approche usager ont été antérieurement regroupées sous l'expression *Information needs and uses*, mais des distinctions sont progressivement apparues, aboutissant à l'émergence de nouvelles définitions. L'*Information Behavior* est défini comme l'ensemble des activités humaines en relation avec les sources et canaux d'information, et inclut la recherche « active » (recherche d'information dans des fonds documentaires, par exemple) et « passive » d'information (télévision par exemple). L'*Information Seeking Behavior*, qui fait également partie des descripteurs MeSH, est défini comme l'activité de recherche intentionnelle de l'information, comme conséquence d'un besoin à satisfaire. A cette fin, les individus interagissent avec des systèmes d'information manuels et/ou informatisés. L'*Information Searching Behavior* est le « niveau micro » de l'interaction avec l'ensemble des systèmes d'information : sont concernées toutes les actions physiques (clics de souris par exemple), mais aussi les opérations cognitives comme les jugements de pertinence. Certains auteurs considèrent que cette activité (*Information Searching Behavior*) est à considérer comme un sous ensemble de la recherche d'information (*Information Seeking Behavior*), elle-même s'intégrant dans le cadre plus large de l'activité informationnelle (*Information Behavior*) (8).

1.1.3. Le besoin d'information

Cette notion est centrale dans le domaine de la recherche d'information, puisque définie comme une interaction entre « un individu qui a besoin d'information » et « un document qui contient ou non la réponse à ce besoin ».

Le besoin d'information a été défini au cours des années comme :

- Rechercher une réponse à sa question. Le besoin est ici défini comme un processus cognitif à quatre niveaux : le besoin réel mais inexprimable,

vaguement conscient, le besoin conscient, mais toujours inexprimable, le besoin pouvant être exprimé en langage naturel, qui a pris la forme d'une question, et enfin le besoin adapté, où la question est présentée au système d'information dans un langage de compromis (mots-clés). Ce modèle a le mérite de mettre en avant la difficile prise de conscience du besoin d'information mais les modalités de passage d'un niveau à un autre ne sont pas envisagées (8).

- Réduire l'incertitude ou un niveau de connaissance insatisfaisant. Cette façon de présenter le problème présente l'intérêt d'introduire l'idée de doute, d'incertitude. Ici, l'incertitude ne doit pas être définie comme un manque de connaissances, mais comme la prise de conscience d'un manque de connaissances (8).
- Donner du sens. Ici, l'information n'est pas une donnée qu'il suffit d'extraire, mais une création personnelle de sens ancrée à un moment donné (8).

1.2. Médecine générale et recherche d'information en santé en ligne

1.2.1. La médecine fondée sur les preuves

La médecine fondée sur les preuves est née à l'Université de McMaster au Canada au début des années 1980. C'était à l'origine un outil pédagogique centré sur l'auto-apprentissage d'un petit groupe d'élèves encadrés par un tuteur (9). En 1996, Sackett va adapter ce principe à la pratique médicale (10). Il le définit comme le fait « *d'utiliser de manière rigoureuse, explicite et judicieuse les preuves actuelles les plus pertinentes lors de la prise de décisions concernant les soins à prodiguer à chaque patient* ». Ce concept a marqué un changement radical dans la pratique médicale, en imposant une rigueur à travers une démarche scientifique et rationalisée dans les prises de décision. Il s'agit de prendre des décisions intégrant les données actuelles de la science et l'expertise clinique du médecin, mais ces décisions doivent être adaptées aux besoins et préférences du patient (Figure 1).

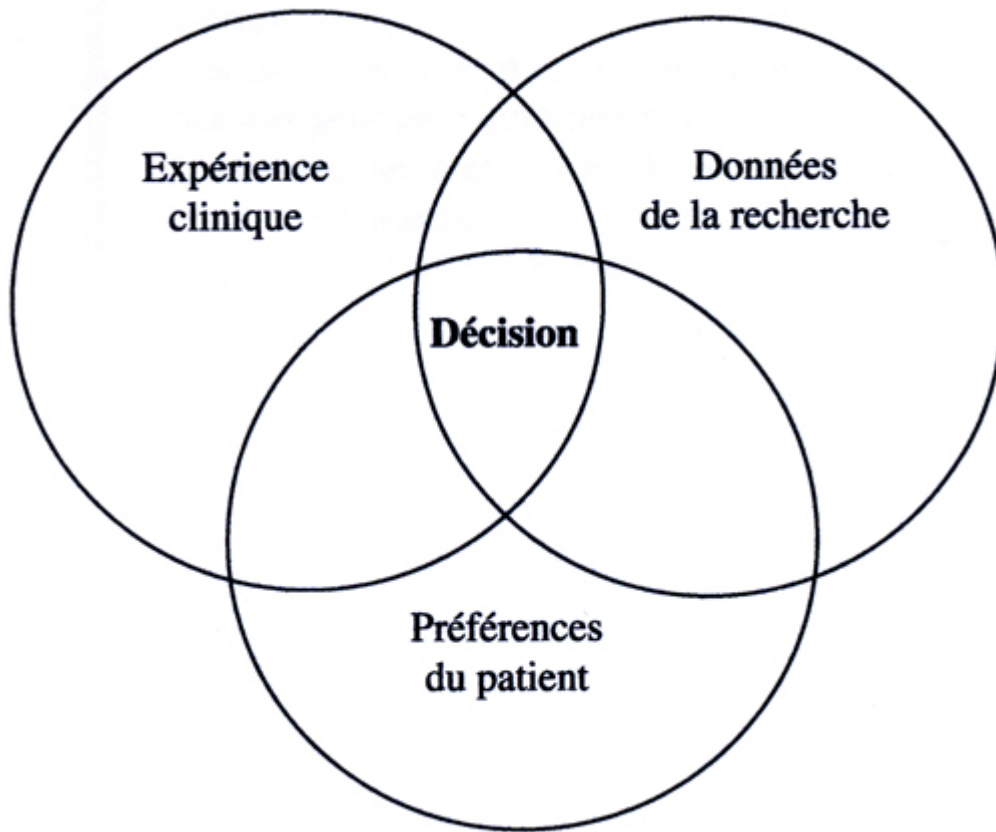


Figure 1. Représentation de la décision médicale basée sur la médecine fondée sur les preuves. D'après Sackett et al., 1996.

La pratique de la médecine fondée sur les preuves nécessite plusieurs étapes (11) :

- La formulation claire et précise d'une question à partir d'un problème donné,
- La recherche d'articles pertinents dans la littérature,
- L'évaluation systématique de la validité et de l'intérêt des résultats et l'extraction des preuves qui sont à la base des décisions cliniques,
- L'intégration de ces preuves dans la pratique médicale courante afin de répondre à la question initiale, en tenant compte des besoins et préférences du patient,
- L'évaluation du processus décisionnel et son ajustement si besoin.

La problématique de la recherche d'information concerne au premier lieu la seconde étape de cette liste, mais elle constitue la colonne vertébrale de l'ensemble de la procédure décrite.

1.2.2. Les besoins en information des médecins généralistes

1.2.2.1. Analyse quantitative des informations utilisées et des questions posées

Le champ d'exercice de la médecine générale, de par sa taille et sa variété, impose à ses praticiens de prêter une attention rigoureuse et continue à la gestion des connaissances. Une étude états-unienne basée sur l'analyse du vocabulaire diagnostique de 11 médecins généralistes au cours de quatre années d'exercice a été effectuée à la fin des années 1970. Pour 1000 patients pris en charge au cours d'une année, chaque praticien rencontrait près de 500 situations cliniques différentes (12). Des praticiens expérimentés utilisent en moyenne deux millions de données pour prendre en charge leurs patients (13).

L'analyse du comportement de 30 médecins généralistes au cours de leur exercice montrait que les praticiens se posaient en moyenne une question clinique tous les 15 patients (14). Les médecins exerçant en milieu rural et ceux ayant une activité plus importante avaient tendance à se poser moins de questions que leurs confrères. Plusieurs enquêtes plus récentes ont également été effectuées auprès de médecins généralistes français (15)(16). Elles retrouvaient toutes deux des résultats similaires. Le nombre moyen quotidien de questions ne trouvant pas immédiatement de réponse de par les connaissances propres du praticien était d'une question pour huit patients vus.

1.2.2.2. Analyse qualitative des questions posées

A la fin des années 1990, Ely s'est attaché à la création d'une taxonomie des questions cliniques posées par des médecins généralistes prenant en charge leurs patients. Dans un premier temps, une étude observationnelle a été menée auprès d'un échantillon de 103 médecins généralistes américains (17). Les investigateurs se sont rendus chez les praticiens pendant deux demi-journées et ont recueilli leurs questions, des plus précises (« *Quelle est la dose de metformine à introduire chez ce patient ?* ») aux plus vagues (« *Quel est ce rash cutané chez cet enfant ?* »). Les participants ont posé un total de 1101 questions, catégorisées au sein d'une taxonomie de questions génériques. Cette taxonomie a ensuite été validée à l'aide d'un second corpus de questions cliniques recueillies auprès de 49 autres médecins généralistes (18). Les trois questions génériques les plus fréquentes étaient « *Quel est le médicament indiqué pour la maladie X ?* » (11%), « *Quelle est la cause du symptôme X ?* » (8%) et « *Quel est l'examen indiqué dans la situation X ?* » (8%). Les dix

questions génériques les plus fréquentes sont présentées dans le tableau 1, elles représentaient 63% de l'ensemble des questions recueillies.

Question générique	Fréquence (%)
Quel est le médicament indiqué pour la maladie X ?	11
Quelle est la cause du symptôme X ?	8
Quel est l'examen indiqué dans la situation X ?	8
Quelle est la dose du médicament X ?	7
Comment dois-je traiter la maladie X ? (Sans limitation au traitement médicamenteux)	6
Comment dois-je prendre en charge la maladie X ? (Prise en charge diagnostique ou thérapeutique)	5
Quelle est la cause de ce signe physique ?	5
Quelle est la cause du résultat de cet examen ?	5
Le médicament X peut-il causer l'effet (indésirable) Y ?	4
Ce patient peut-il avoir la maladie X ?	4

Tableau 1. Questions génériques issues des questions posées par les médecins généralistes et leurs fréquences. D'après Ely, 2000.

Comme évoqué par l'auteur, les perspectives potentielles de ces résultats étaient multiples. Ils pouvaient permettre de développer des outils destinés à relier questions cliniques et réponses. Cette taxonomie pouvait également être utilisée pour identifier et classer les domaines où les sources d'information pêchent à répondre à certaines questions cliniques. Enfin, elle pourrait également servir à identifier des priorités de recherche, notamment dans les domaines où les réponses sont insuffisantes ou inexistantes.

En 2007, une étude menée à Madrid auprès de médecins généralistes a permis de recueillir 635 questions au cours de 3511 consultations (19). Ces questions ont été catégorisées à l'aide de la taxonomie proposée par Ely. Les questions les plus fréquentes se rapportaient au diagnostic (53%) et au traitement (26%) des maladies. Moins de 2% des questions se référaient à des éléments relatifs à la communication entre le médecin et le patient. Les questions non-cliniques représentaient 13% de l'ensemble des questions, elles concernaient essentiellement des problématiques administratives et de formation médicale continue. Ces résultats sont confirmés par une revue de la littérature publiée en

2010 : les besoins d'information des médecins généralistes concernent la prise en charge clinique des patients, la formation médicale continue et l'information au patient (20).

1.2.3. Les supports d'information en santé disponibles en ligne pour les professionnels de santé

Dans cette partie, nous présentons les principaux supports d'information en santé disponible en ligne pour les professionnels de santé, et notamment à destination des médecins généralistes. Nous avons classé ces différents supports sur la base du rapport concernant les sources d'information et la méthodologie de recherche documentaire rédigé par l'Institut de Recherche et de Documentation en Economie de la Santé (IRDES) actualisé en avril 2016, et présenterons pour chacune de ces catégories les principaux supports d'information (21).

Le rapport de l'IRDES permet d'identifier différents supports d'information :

- Les bases de données bibliographiques,
- Les agences gouvernementales,
- Les sociétés savantes, universités médicales et revues médicales,
- Les portails ou répertoires,
- La littérature grise, même s'il paraît plus adapté de la classer parmi les sources d'information et non les supports.

A cette liste, nous pouvons ajouter :

- Les bases de données médicamenteuses,
- Les outils d'aide à la décision,
- Les sites d'associations de patients, reconnues par la Haute Autorité de Santé comme catégorie à part entière dans son guide sur la recherche d'informations médicales sur Internet (22),
- Les sites destinés au grand public,
- Les applications mobiles,
- Les médias sociaux.

1.2.3.1. Bases de données bibliographiques

Une base de données est « un ensemble de données relatif à un domaine défini de connaissances et organisé pour être offert aux utilisateurs. » (23). La base de données

bibliographiques est un ensemble structuré de données sous forme de références, accessibles au moyen d'un logiciel (24).

La base MEDLINE, créée par la National Library of Medicine (NLM) est la plus grande base de données bibliographiques en santé. Elle regroupe plus de 27 millions d'articles publiés dans plus de 5000 revues, dans tous les domaines du champ biomédical. Tous les résumés, ainsi que la grande majorité des articles, sont rédigés en anglais. Elle est accessible par le moteur de recherche PubMed, via l'utilisation du thésaurus MeSH (25). D'autres bases de données bibliographiques gratuites (comme Google Scholar) ou payantes (comme Embase) sont également disponibles (26)(27).

La librairie Cochrane met à disposition des revues systématiques de la littérature, régulièrement actualisées. Elles sont réalisées par la Cochrane Collaboration, réseau international de correspondants chargés d'analyser systématiquement les essais contrôlés randomisés réalisés dans diverses spécialités médicales (28).

Dans le domaine francophone, le catalogue CISMef (Catalogue et Index des Sites Médicaux de langue Française) recense 85 bases de données bibliographiques concernant des domaines variés allant de la santé au travail au handicap, en passant par l'éducation à la santé ou la psychiatrie (29). Parmi les plus consultés, citons la Banque de Données Santé Publique (BDSP), réseau de coopération documentaire auquel participent plusieurs dizaines d'institutions françaises spécialisées dans le champ de la santé publique (30).

1.2.3.2. Sites d'agences gouvernementales

Dans le domaine francophone, le catalogue CISMef recense 31 sites d'agences gouvernementales, dont 19 issus d'agences françaises. Parmi celles-ci, nous pouvons citer la Haute Autorité de Santé, l'Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM) et l'Agence Santé Publique France, qui résulte de l'union récente de l'Institut de Veille Sanitaire (InVS), l'Institut National de Prévention et d'Éducation à la Santé (INPES) et l'Établissement de Préparation et de Réponse aux Urgences Sanitaires (EPRUS) (31)(32).

1.2.3.3. Sites de sociétés savantes, universités médicales et revues médicales

Dans le domaine francophone, le catalogue CISMef recense 332 sites de sociétés savantes, dont 188 françaises. Dans ce même catalogue sont indexés 637 périodiques concernant le

domaine de la santé, dont 500 français. Plus de 5000 périodiques sont indexés dans MEDLINE.

De nombreuses universités médicales françaises se sont impliquées dans la diffusion de connaissances dans le domaine médical. Citons notamment les universités de Strasbourg et de Grenoble. En plus des enseignements classiques, les établissements de l'enseignement supérieur développent depuis quelques années des MOOC (*Massive Online Open Course*, formations en ligne ouvertes à tous), qui sont des formations à distance, en ligne, ouvertes, capables d'accueillir un grand nombre de participants. La plateforme France Université Numérique propose actuellement 35 MOOC gratuits dans le domaine de la santé (33). D'autres formations de ce type sont également proposées par des entreprises ou des structures privées.

Depuis 2003, les universités françaises ayant des composantes en médecine adhèrent à l'Université Médicale Virtuelle Francophone (UMVF), groupement d'intérêt public cofinancé par les universités et des subventions du Ministère de l'Enseignement Supérieur et de la Recherche. Il devient en 2009 l'Université Numérique Francophone des Sciences de la Santé et du Sport (UNF3S) en intégrant les disciplines pharmacie, odontologie et sciences du sport (34). L'UNF3S propose, via ses sites internet, en accès libre et gratuit, des ressources pédagogiques validées par les collèges de spécialités de ses composantes ou issues des sites des facultés françaises, par spécialité et année d'étude.

1.2.3.4. Portails et catalogues

Un portail documentaire constitue un point d'accès unique à de multiples services et ressources documentaires électroniques (35). Un catalogue est une base de données signalant tout ou partie des documents disponibles dans une ou plusieurs bibliothèques, il est destiné à identifier les documents et faciliter leur recherche ou leur localisation (36). SUDOC est le catalogue du Service Universitaire de Documentation. C'est le catalogue collectif français réalisé par les bibliothèques et centres de documentation des établissements de l'enseignement supérieur et de la recherche. Il a notamment pour fonction de recenser l'ensemble des thèses produites en France (37).

Orphanet est un portail d'informations sur les maladies rares et les médicaments orphelins, créé par l'INSERM en 1997. Il propose notamment un inventaire et une classification des maladies rares, une encyclopédie sur les maladies rares et un répertoire

des ressources expertes (38). A la différence de SUDOC ou de CISMeF, il associe à sa fonction de portail documentaire une fonction de production de données.

CISMeF est un projet initié par le Centre Hospitalier Universitaire de Rouen ayant débuté en 1995 (24). Ce catalogue^b indexe les principaux sites et documents francophones dans le domaine de la santé. Il utilise essentiellement deux outils standards pour organiser l'information : le thésaurus MeSH (*Medical Subject Heading*) et le format de métadonnées du Dublin Core (39). Il recense actuellement près de 118 000 ressources. Il est dévolu aux professionnels de santé, mais également au grand public. Il comporte trois axes principaux : la médecine factuelle, l'enseignement et l'information aux patients. Il constitue aujourd'hui le site-catalogue de référence dans la santé francophone.

1.2.3.5. Bases de données médicamenteuses

Parmi elles, on peut notamment citer la base VIDAL, base historique dont la création remonte au début du XX^{ème} siècle, la base Claude Bernard, qui fut la première base de données électronique sur les médicaments agréée par la HAS dans le cadre de la certification des logiciels d'aide à la prescription et la base Thériaque issue du centre national hospitalier d'information sur le médicament (40)(41)(42). Les bases Thesorimed, développée par l'Assurance Maladie et Clickadoc, développée par le centre de documentation de l'Office Commercial Pharmaceutique (OCP) viennent compléter la liste des bases de données médicamenteuses agréées par la HAS (43)(44).

De création plus récente, la base de données publique des médicaments permet au grand public et aux professionnels de santé d'accéder à des données et documents de référence sur les médicaments commercialisés ou ayant été commercialisés dans les trois dernières années en France (45). Cette base de données administratives et scientifiques sur les traitements et le bon usage des produits de santé est mise en œuvre par l'ANSM, en liaison avec la HAS et l'Union Nationale des Caisses d'Assurance Maladie (UNCAM), sous l'égide du Ministère des affaires sociales et de la santé.

^b En dépit du terme de « catalogue » de son acronyme et du caractère très divers des ressources qu'il propose, on peut aussi considérer cet outil comme une base de données bibliographiques.

1.2.3.6. Outils d'aide à la décision

Les systèmes d'aide à la décision médicale sont définis comme des outils informatiques « dont le but est de fournir aux cliniciens en temps et lieux utiles les informations décrivant la situation clinique d'un patient ainsi que les connaissances appropriées à cette situation, correctement filtrées et présentées afin d'améliorer la qualité des soins et la santé des patients. » (46). Il en existe pour l'ensemble des activités médicales (prévention, dépistage, diagnostic, traitement), les différentes catégories de médecins (spécialistes, généralistes, étudiants) et de modalités d'exercice (libéral ou hospitalier) (47). Plus récemment, des systèmes d'aide à la décision ont été développés à destination des patients en vue d'une décision partagée.

En dehors des logiciels d'aide à la prescription, destinés à la sécurisation de l'ordonnance, on peut classer ces outils selon qu'ils aident à la décision diagnostique ou thérapeutique. On peut aussi différencier les outils selon qu'ils utilisent une approche numérique fondée sur les données ou une approche symbolique fondée sur les connaissances.

Les outils d'aide à la décision basés sur des données peuvent utiliser des approches probabilistes pour l'aide au diagnostic (comme l'outil visualDx, par exemple), des modèles pronostiques pour l'aide à la thérapeutique (comme le site <https://statindecisionaid.mayoclinic.org/> permettant d'évaluer l'impact en termes d'évènements cardiovasculaires de l'ajout d'une statine chez un patient, en fonction de son risque cardiovasculaire global) ou permettre le calcul de scores (comme le Mini Mental State Examination dans l'évaluation des troubles cognitifs) (48)(49).

Les outils d'aide à la décision fondés sur les connaissances peuvent être basés sur des systèmes experts. Ceux-ci sont classiquement constitués de trois modules : une base de connaissances contenant les connaissances médicales théoriques et empiriques nécessaires à la démarche diagnostique ou thérapeutique, dans un domaine d'expertise, une base de faits représentant les caractéristiques du patient pour lequel on cherche à déterminer le diagnostic ou la prise en charge thérapeutique, un moteur d'inférences articulant les connaissances de la base de connaissances pour raisonner sur la résolution du problème posé dans la base de faits. Ils peuvent également se baser sur des systèmes de mise en œuvre des recommandations de pratique clinique. Depuis le début des années 2000, des approches qualifiées de documentaires ont été utilisées pour permettre aux cliniciens d'accéder au contenu des recommandations de bonne pratique de manière ciblée dans une démarche de recherche d'information. Parmi ces outils, on peut

citer UpToDate et ClinicalKey, services payants proposant des fiches d'aide au diagnostic, à la thérapeutique et à l'éducation du patient ou VidalRecos, recueil de recommandations développé par VIDAL et qui propose des informations textuelles et des arbres décisionnels sous forme de schéma (50)(51)(52)(47). Depuis quelques années, plusieurs outils d'aide à la décision thématiques, à destination essentiellement des médecins généralistes, ont été mis en ligne. La plupart d'entre eux a vu le jour sous l'impulsion d'un département universitaire de médecine générale. On peut citer Antibioclic pour la prise en charge des infections bactériennes, Gestaclic pour le suivi des femmes enceintes, Pédiadoc pour le suivi du développement de l'enfant ou Aporose pour la prise en charge de l'ostéoporose (53)(54)(55)(56).

1.2.3.7. Sites d'associations de patients

Plus de 2000 associations de patients dans le monde francophone sont référencées sur CISMeF, dont près des trois quarts en France. Elles sont généralement centrées autour d'une pathologie ou d'un handicap. La très grande majorité de ces associations dispose d'un site internet, où des informations sur la maladie ou le handicap en question sont parfois disponibles.

1.2.3.8. Sites destinés au grand public

Ils regroupent les encyclopédies en ligne, dont la plus connue est Wikipédia, les forums de santé et tous les autres sites fournissant des informations sur la santé à destination du grand public. Parmi ces sites, on distingue les sites institutionnels, les sites marchands et les sites d'information grand public, la frontière entre ces deux dernières catégories demeurant souvent floue. Les sites institutionnels (Améli, Santé Publique France) sont nombreux et accessibles à tous, ils diffusent une information publique, produite et contrôlée par les services publics (57)(32). Les sites marchands peuvent être créés par des assureurs, des producteurs de dispositifs médicaux ou de médicaments. Ils produisent une information en lien direct ou indirect avec les produits qu'ils commercialisent. Les sites d'information grand public poursuivent à la fois des objectifs d'information, de prévention, d'échange mais également des objectifs marchands. Ils comportent en général une partie rédactionnelle et des espaces communautaires (58).

1.2.3.9. Applications mobiles

En 2016, on dénombrait 165 000 applications mobiles dans le domaine de la santé, dont près de la moitié ont été téléchargés plus de 10 000 fois. Environ 15% de ces applications étaient destinées aux médecins (59). Cette même année, le baromètre numérique du CESSIM (Centre d'Etudes sur les Supports de l'Information Médicale) montrait que 50% des médecins généralistes avaient utilisé au moins une application au cours des 30 derniers jours sur smartphone ou sur tablette (60). Parmi les médecins ayant un smartphone, 58% utilisent des applications médicales, proportion stable depuis 2012. Les applications les plus utilisées sont les bases de données médicamenteuses (88% des utilisateurs), suivies des applications dédiées aux interactions médicamenteuses (46% des utilisateurs). D'après les médecins, les sociétés savantes et autorités de santé sont les mieux placées pour leur recommander une application mobile (61). Parmi les acteurs du secteur, citons la société DMD Santé, crée par le psychiatre Guillaume Marchand, dont l'objectif est d'évaluer la qualité des applications en santé disponibles en français.

Dans un champ proche de celui des applications mobiles, citons également les jeux sérieux (*serious games*), dispositifs pédagogiques scénarisant un processus de formation visant à l'acquisition de compétences opérationnelles. Ils peuvent être utilisés en présentiel ou à distance. Ils se développent beaucoup dans le domaine de la santé, où ils concernent aussi bien la formation des personnels de santé que celle des patients et aidants (62). Dédié aux cardiologues et aux urgentistes, le jeu SIMUrgences est un module de jeu 3D en temps réel qui forme les médecins à la prise en charge des patients en situation d'urgence cardiologique (63).

1.2.3.10. Médias sociaux

L'Association Internationale d'Informatique Médicale a classé les différents médias sociaux en 13 catégories différentes. Parmi celles-ci, on trouve notamment les réseaux sociaux (comme Facebook), les blogs et microblogs (comme Twitter), les wikis, les forums et listes de diffusion et les applications sociales (64). L'apparition des médias sociaux a permis l'accroissement exponentiel des possibilités de communication entre professionnels de santé. Elle a facilité le réseautage professionnel, via le connectivisme, théorie de l'apprentissage basée sur les apports des nouvelles technologies, et la création de communautés de pratiques (65).

2. Les comportements des médecins généralistes à la recherche d'informations en santé en ligne

2.1. De la question à la recherche d'une réponse

Au cours des années 2000, plusieurs études ont analysé le comportement des médecins généralistes à la recherche d'informations en ligne. La proportion de questions posées pour lesquelles une réponse était cherchée par le praticien variait selon les études. Elle passait de 23% (19) à 57% (66) voire 61% dans une étude menée en 2008 auprès de médecins généralistes français (16). Ils tentaient de répondre à 42% de ces questions pendant la consultation et à 58% après la consultation (16). Les questions posées pendant la consultation trouvaient plus de réponses satisfaisantes que les questions posées après la consultation (100% versus 75%) (19). Une étude menée par Ely en 2007 auprès de médecins généralistes retrouvait un taux de réponse aux questions posées de 59% (67). Une revue de la littérature publiée en 2014 retrouvait un taux de réponse de 78%, mais concernait l'ensemble des professionnels de santé (68).

Au Canada, dans une étude parue en 2013, les généralistes interrogés utilisaient préférentiellement les livres et l'avis d'un confrère en cas de questions clinique. Pour leur formation continue, Internet arrivait en 3^{ème} position avec 49% d'utilisation, derrière les revues médicales (93%) et les guides de pratique clinique (70%) (69). Les sources d'informations préférées des médecins généralistes français étaient, par ordre de préférence : les congrès et sessions de formation continue, Internet et les livres et revues imprimées (70).

2.2. Les supports d'information en ligne des médecins généralistes

Plusieurs études se sont intéressées aux supports d'information en ligne des médecins généralistes. Elles sont majoritairement limitées par leur caractère déclaratif. Une étude effectuée auprès de 300 généralistes néo-zélandais montrait que, parmi les utilisateurs d'Internet, MEDLINE constituait le principal support d'information (71). Une récente étude écossaise montrait que, sur les 37 supports d'information en ligne citées par les généralistes interrogés, les plus utilisés étaient le site des recommandations des collèges médicaux écossais (SIGN Guidelines, 79,3%), le site de formation continue du British Medical Journal (BMJ Learning, 73%), le site des recommandations du NICE (National Institute for Health and Clinical Excellence, 66,8%), une encyclopédie médicale en ligne (GP Notebook, 63,6%) et Google (61,3%) (72). Les réseaux sociaux (Facebook et Twitter) restaient très marginalement utilisés.

En France, une analyse des historiques de navigation d'un échantillon de médecins généralistes a été menée en 2015 (73). Les 10 sites ou catégories de sites les plus consultés étaient les suivants :

- Le site de VIDAL (25,7%)
- Des sites grand public (8,6%), comme Wikipédia ou des forums en lien avec la santé,
- Des sites de revues médicales (7,3%),
- Le site de l'Assurance Maladie (7%),
- Des sites d'hôpitaux (6%),
- Des sites d'universités (5,1%),
- Des sites d'institutions publiques en santé (4,8%), comme la Haute Autorité de Santé, l'Institut National de Veille Sanitaire ou l'Institut National de Promotion et d'Education en Santé,
- Le site du CRAT (4,8%),
- Des sites de formation médicale financés par l'industrie pharmaceutique (4,1%), comme Univadis,

- Des sites institutionnels (3,8%), comme le site du Conseil National de l'Ordre des Médecins ou celui du Ministère de la Santé.

Une récente revue de la littérature s'est penchée sur la manière dont les professionnels de santé utilisaient les médias sociaux pour développer des communautés virtuelles facilitant le réseautage professionnel, le partage de connaissances et une pratique basée sur des données validées (65). Elle montre que les professionnels de santé conservent une utilisation limitée des médias sociaux à des fins professionnelles. Les utilisateurs semblent plus enclins à intégrer des communautés spécifiques de leur discipline ou de leur champ d'exercice. Au sein de ces communautés, la plupart des contributions sont attribuées à un petit nombre d'utilisateurs et les membres observateurs justifient leur appartenance à la communauté par l'accès potentiel à des informations importantes. Même s'il persiste un scepticisme sur la validité de l'information, certains médias, comme Twitter, pourraient jouer un rôle dans la réduction du fossé entre données validées et pratique clinique (65).

2.3. Les obstacles à la recherche d'information

En 2002, Ely identifiait près de 60 obstacles à la recherche de données validées en ligne par des médecins généralistes (74). Parmi eux, six semblaient particulièrement saillants :

- Le temps nécessaire à la recherche d'informations,
- Les difficultés à reformuler la question initiale, souvent vague et pouvant être interprétée de multiples façons,
- Les difficultés à identifier une stratégie de recherche optimale,
- L'absence de sources valides d'information,
- L'incertitude quant au fait que toute l'information pertinente a été retrouvée et que la recherche peut être stoppée,
- L'absence d'une approche synthétique des données validées permettant une réponse clinique pertinente pour le patient.

Une synthèse de la littérature effectuée plus récemment (70) classait les obstacles à la recherche d'information en ligne par les médecins généralistes en quatre catégories :

- Les obstacles liés aux médecins : préférence pour d'autres sources d'information, manque de compétences à la recherche d'information en ligne, âge élevé,
- Les obstacles liés aux conditions d'exercice : manque de temps, forte charge de travail, absence de valorisation du temps passé à la recherche d'information, risque d'altération de la relation entre le médecin et le patient,
- Les obstacles liés à la technologie : difficultés de connexion du fait de débits trop faibles, coût des connexions, gestion des codes d'accès et des mots de passe, crainte des virus, manque de convivialité des supports d'information en ligne,
- Les obstacles liés aux informations en ligne : difficultés à gérer le « trop-plein » d'information, absence de maîtrise de l'anglais, inquiétudes concernant la qualité et la fiabilité des informations, manque de pertinence des informations.

2.4. Attitudes et comportements des internes et médecins généralistes à la recherche d'information sur Internet

2.4.1. Contexte

Les médecins généralistes sont encouragés à pratiquer une médecine fondée sur les preuves ou *Evidence-Based Medicine* (EBM). Ils font face à une grande variété de patients et de situations cliniques durant leur exercice. Cela génère de nombreuses questions liées à la prise en charge des patients. Internet rend disponible un large spectre de réponses possibles à ces questions, et est de plus en plus utilisé par les médecins généralistes. Cependant, de nombreuses questions restent sans réponse, mettant en évidence un fossé entre les besoins d'information des médecins et les informations disponibles en ligne. Plusieurs barrières à la recherche d'information en ligne par les médecins généralistes ont déjà été identifiées dans la littérature. Ces freins déjà identifiés suffisent-ils à expliquer l'absence de congruence entre besoins d'information et pratiques réelles des médecins généralistes ? Il nous a semblé qu'une étude qualitative effectuée auprès de médecins généralistes et d'internes en médecine générale nous aiderait à répondre à cette question.

2.4.2. Objectif

Comprendre les comportements et attitudes des généralistes et des internes de médecine générale dans la recherche en ligne d'informations.

2.4.3. Méthode

Etude qualitative par entretiens collectifs semi-dirigés, avec analyse thématique des données.

Cinq entretiens collectifs ont été menés entre octobre 2013 et janvier 2014, auprès de 20 généralistes (en Haute-Normandie ou Ile de France) et de 15 internes (de Haute-Normandie). Pour les médecins, l'échantillon a été raisonné sur le genre, le type d'exercice (libéral/remplaçant), le lieu d'exercice (rural/urbain), le nombre d'années de pratique de la médecine générale et le niveau d'informatisation du cabinet. Pour les internes,

l'échantillon a été raisonné sur le nombre de stages ambulatoires en médecine générale. L'échantillon total a aussi été raisonné sur le l'âge et le sexe. Lors de chaque entretien collectif, un modérateur décrivait brièvement l'objectif de l'étude, animait la séance, vérifiait que tous les items du guide étaient abordés et assurait la participation de tous les volontaires. Simultanément, un observateur notait la communication non verbale et les interactions entre participants. Les données ont été recueillies par enregistrement audio. La transcription des verbatim a été anonymisée avant le codage. Celui-ci s'est déroulé en 2 étapes. Un premier codage, ouvert, a été effectué de façon indépendante par deux auteurs. L'équipe a ensuite discuté ensemble les codes et les a regroupés en catégories. L'analyse a été réalisée avec le logiciel NVivo 10®. Cette analyse s'est déroulée en parallèle du recueil des données. La saturation a été obtenue après cinq entretiens.

2.4.4. Résultats

Les sujets faisant l'objet de recherche sur Internet étaient les problèmes de santé les plus fréquents en soins primaires mais aussi la prévention, l'approche centrée patient, la communication, la coordination des soins ou le professionnalisme. Les généralistes se servaient d'un nombre limité de sites. Ils considéraient ces sites comme fiables, sans se servir nécessairement de critères d'évaluation. Quand ils avaient des critères de sélection, ceux-ci étaient : l'origine institutionnelle du site, l'absence de financement par l'industrie, l'apparence du site, la réactualisation et le référencement des informations. Pour les autres, ils s'estimaient soit incompetents soit illegitimes pour évaluer la fiabilité des sites consultés. Les informations cherchées étaient directement en lien avec leur pratique clinique. Les stratégies de recherche en ligne étaient variables. Elles se faisaient le plus souvent dans le langage naturel, les participants étant peu sensibilisés à l'emploi des mots-clés et d'opérateurs booléens. Google® était le moteur de recherche le plus employé. Les participants étaient conscients des conséquences de la qualité de la stratégie de recherche sur la validité des données trouvées. Ils se servaient de requêtes rapides sur leur lieu de travail et de requêtes plus larges à domicile. Ils citaient l'anglais et l'accès payant de certains sites comme des obstacles à la recherche en ligne.

Une présentation détaillée de la méthode et des résultats de cette étude figure ci-après.



Behavior and attitudes of residents and general practitioners in searching for health information: From intention to practice



Matthieu Schuers^{a,b,*}, Nicolas Griffon^{a,c}, Gaëtan Kerdelhue^a, Quentin Foubert^b, Alain Mercier^d, Stéfan J. Darmoni^{a,c}

^a CISMef, TIBS, LITIS EA 4108, Rouen University Hospital, rue de Germont, Rouen, France

^b Department of General Medicine, Rouen University, 22 boulevard Gambetta, Rouen, France

^c UMR.S 1142, LIMICS, Inserm, Paris, France

^d Department of General Medicine, Paris 13 University, 74 rue Marcel Cachin, Bobigny, France

ARTICLE INFO

Article history:

Received 13 October 2015

Received in revised form 9 February 2016

Accepted 12 February 2016

Keywords:

General practice

Information seeking behavior

Internet

ABSTRACT

Background: Physicians are increasingly encouraged to practice evidence-based medicine (EBM), and their decisions require evidence based on valid research. Existing literature shows a mismatch between general practitioners' (GPs) information needs and evidence available online. The aim of this study was to explore the attitudes and behavior of residents in general medicine and GPs when seeking medical information online.

Methods: Five focus groups (FGs) involving residents in general medicine and GPs were conducted between October 2013 and January 2014. The overall number of participants recruited was 35. The focus group discussion guide focused on participants' experiences in searching for health information on the Internet, perceived barriers and possible solutions for improving the quality of their own search processes. Descriptive analysis was performed by three researchers.

Results: Participants described a wide range of research topics, covering all general medicine core competencies, and especially patient-oriented topics. They used a limited list of websites. Participants were not confident about their ability to assess the quality of the information they found. Their assessment of data quality was based on intuition, and they mainly sought concordance with their existing knowledge. The way the data were exposed was considered very important. Participants were looking for information that was directly linked to their clinical practice. Information seeking processes varied among participants. They felt they had not mastered query building for conducting searches, and were aware of the impact this shortcoming had on the quality of their search for information.

Conclusions: Residents in general medicine and GPs understood the importance of EBM and the need for objective and reliable information. The present study highlights the difficulties in identifying this kind of information, and suggests ideas for improvement. Available search tools should change in order to fill the gap with real-world clinical practice, for example by integrating a patient-centred approach.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Background

Evidence-based medicine (EBM) is a combination of individual clinical expertise and best available external evidence alongside patient's values and expectations [1]. Currently, EBM is increasingly encouraged in physicians' practice, and decision-making, requiring evidence based on valid research.

General practitioners (GPs) face a wide range of patients and clinical situations during their practice. This generates a large number of questions on patient management, varying between 0.10 and 1.85 questions per patient [2]. The Internet provides a broad spectrum of possible answers to GPs' questions, and is increasingly used by physicians in their daily practice [3,4]. Nevertheless, despite the

Abbreviations: EBM, evidence based medicine; FG, focus group; GP, general practitioner.

* Corresponding author at: Department of General Medicine, Rouen University, 22 boulevard Gambetta, Rouen, France.

E-mail addresses: matthieu.schuers@gmail.com (M. Schuers), nicolas.griffon@chu-rouen.fr (N. Griffon), gaetan.kerdelhue@chu-rouen.fr (G. Kerdelhue), quentin.foubert@gmail.com (Q. Foubert), alain.mercier@univ-paris13.fr (A. Mercier), stefan.darmoni@chu-rouen.fr (S.J. Darmoni).

growing availability and use of online information sources, some clinical questions remain unanswered [5]. This clearly shows a mismatch between GPs' information needs and evidence available online.

Although previous studies have tried to identify the information needs of GPs, most were involving north-American physicians, making the results difficult to duplicate in other practice contexts [6]. Based on the findings of these studies, GPs' information needs mostly concern treatments, diagnosis and drug therapies. Patient-centred topics, such as education or communication skills, remain rarely cited. Some barriers for the use of Internet for information seeking have already been identified. Time constraints are commonly reported in literature [2]. The lack of skills to perform a literature search, information overload and heterogeneity in information quality are frequently cited [3,7,8].

But are these known barriers sufficient to explain the remaining gap between intention and practice in finding health information online? Most of existing data come from quantitative studies, and we strongly believe that qualitative data could help us to deeply understand how residents and GPs search for medical information online, their research topics and especially the difficulties they encounter when seeking information.

The aim of this present study was to describe and understand the attitudes and behaviors of residents in general medicine and GPs in two regions of northern France regarding online information searching. More precisely, we aimed to identify the medical topics sought, any barriers encountered in retrieving information, and the methods for overcoming such obstacles.

2. Methods

2.1. Design

We conducted a qualitative study. Data were collected through 5 focus groups (FG) interviews. We recruited residents in general medicine and GPs to join the focus groups. Dynamic interaction among the participants motivated our choice of data collection technique [9]. Data were analyzed using a qualitative descriptive approach.

2.2. Participant recruitment

Five FGs were conducted between October 2013 and January 2014. Purposive sampling involved recruiting residents with different levels of clinical experience in general practice settings. In France, residency in general medicine lasts 3 years and follows a 6-year undergraduate programme. For the present study, residents were recruited from the second and third year groups of general medicine residents enrolled at the School of Medicine of the University of Rouen, France. Residents were first contacted by email. Sampling was then completed via personal invitations, in order to ensure variability of participants. Physicians were recruited in two regions of northern France (Upper Normandy and the greater Paris area). They were all contacted, by phone, email or face-to-face invitations. They all accepted the invitation. We ensured that FG discussions included younger and more-experienced, male and female private-practice GPs, and locums from both rural and urban areas. We also ensured that their practices had variable levels of computerisation. Some of the physicians recruited were involved in residency teaching. All participants gave their informed consent before participating. This study was approved by the ethical committee of the Rouen University Hospital. Information on participants' profiles is detailed in Section 3.

2.3. Data collection

A semi-structured topic guide was built, based on existing knowledge. The results of the few available quantitative studies focused on research topics and querying processes of physicians, but rarely explored GPs' difficulties when seeking information. The subjects of the focus group discussion guide included participants' experiences in searching for health information on the Internet, perceived barriers and possible solutions for improving the quality of their own process when seeking medical knowledge. Four FG discussions were held outside the physicians' practice, at the Rouen medical school. The fifth one was held in a primary care setting. A moderator facilitated the discussions (MS or AM), briefly explaining the aim of the study, and leading the discussion. The moderator ensured that all issues were covered and that everyone participated. A researcher (QF) took notes during the discussions for further analysis, focusing on nonverbal communication and interactions between participants.

2.4. Analysis

Data analysis occurred concurrently with data collection. Data saturation was achieved after 5 focus groups, so data collection was stopped at this point. All FG discussions were recorded and transcribed verbatim. The transcripts have been de-identified. Confidentiality was ensured by assigning a number to each focus group participant. Only two members of the research team had direct access to the data. When using direct quotations of participants in the presentation of results, we took care to ensure the quotations used did not contain information that may be potentially identifiable. Our aim was to describe and understand the residents and GPs' behavior while searching for health information using internet-based resources. This resulted in a descriptive approach, about the content of websites and the different pathways used to reach this content. Data were processed in different steps, using first open coding. This first phase was performed independently by two researchers (MS and QF), without any predefined framework. The team members then came together to discuss the codes, which were then gathered together into key themes. Codes and emerging themes were compared for coding reliability through a process of discussion and deliberation (MS, QF and AM). N Vivo® software package (version 10) was used to support analysis of the transcripts. The analysis was entirely completed in French. Then, MS and AM translated in English the results and the categories, as well as the quotations included. The quality of the translation was analyzed by a native English speaker, together with the research team.

3. Results

3.1. Participants profile

We recruited 35 participants. Fifteen were residents and 20 were GPs. Mean age was 45 years and 17 were male. The characteristics of the participants and practices are summarized in Table 1.

3.2. Key points

Firstly, a variety of research topics was described. Secondly, the reasons for participants using certain websites when seeking data were identified. Finally their querying habits were described.

3.3. Research topics

Research topics covered all general medicine core competencies. Many queries concerned the most frequent health problems

Table 1
Participants characteristics (residents and GPs, n = 35).

Variables	
Age	45 (26–62)
Male	17 (49%)
Residents (n = 15)	
- Age	27 (26–29)
- Number of 6 month clinical internship in general practice	10 (67%)
- Number of 12 month clinical internship in general practice	5 (33%)
GPs (n = 20)	
- Age of private practice GPs (n = 17)	48 (29–62)
- Age of locum GPs (n = 3)	28 (27–29)
- Male	13 (65%)
- Number of years in practice	22 (1–34)
- Internship instructors	11 (55%)
- University lecturers	10 (50%)
- Internet at GP's practice	16 (80%)

Data are presented as mean (range) or n (%).

encountered in primary care and focused on diagnosis, complementary investigations and treatment issues.

« In April, they changed the child vaccination schedule. We found ourselves groping around in the dark. We needed internet to retrieve the new schedule. » FG1, Female, Resident, 26 y

« If it's a bit complicated, I can easily look up information on care, treatment and follow up. » FG2, Male, Resident, 29 y

« If a patient has an INR at 11, what do I do? I've got the answer in real-time on internet. » FG4, Male, GP, 52 y

Other queries concerned the following competencies:

- Prevention, health education,

« I show pregnant women how to use a website on use of medications during pregnancy. » FG5, Male, GP, 43 y

« What I'm looking for on internet is documents to hand out to patients to help them with problem solving or understanding their health concerns. » FG5, Male, GP, 47 y

- Communication, patient-centred approach,

« Apparently our specialty is the first contact, but in reality it's third: "The specialist suggested such and such, doctor, what do you think about it?". It's terrible, but in fact, it's all on internet! » FG5, Male, GP, 43 y

« I've had problems several times, because I didn't do a PSA test. The patients then saw an urologist, and came back saying: "You have to do the test". Faced with this situation, I've shown them the guidelines on cancer prevention, which concludes that PSA is not a screening test. » FG3, Male, GP, 60 y

« When it's not clear cut, and there is disagreement, we go straight to internet. » FG3, Male, GP, 60 y

- Comprehensive approach, complexity,

« Official guidelines (like the French National Authority for Health: HAS) are too strict, and not practical enough. I think they're sometimes hard to apply in real situations. » FG 4, Female, GP, 45 y

« I'd try to mix information of internet with my own experience and what the patient wants. » FG 4, Male, GP, 52 y

- Coordination of care

« When I refer to my dermatologist I attach a picture and I ask: "What do you think? Do you need to see him/her quickly?" » FG3, Male, GP, 60 y

« My referent radiologist sends me images by internet, which I can directly upload to the patient's file. » FG3, Male, GP, 56 y

- Professionalism

« I also use internet for training purposes, including CME... » FG1, Male, Resident, 28 y

« When I started out, I frequently used ENT images, especially of the eardrum, as it was difficult to identify otitis media. » FG2, Female, Resident, 26 y

« I'm used to going on the ministry of health website, because there are documents on health care organization. » FG4, Male, GP, 58 y

« [With internet], you don't feel so alone, especially if you have support from groups which share their work and papers. » FG5, Male, GP, 43 y

Information seeking objectives seemed to be different for residents and physicians: enabling physicians to update their knowledge, whereas for residents it was a form of self-reassurance.

« I sometimes feel the need to confirm, in the fear of... I always want to be sure that I have the correct information, the correct guideline. » FG 1, Female, Resident, 28 y

« What I'm looking for online, it's documents that will help me to enrich my knowledge, when I need to learn about a specific topic. » FG 5, Male, GP, 47 y

3.4. Categories and concepts relating to Internet resource selection

Participants were more confident using the websites they already knew.

« In fact, I use a website I know well and which I trust. » FG 5, Male, GP, 47 y

In practice, they used a limited list of websites, the content and characteristics of which were familiar to them, thus enabling faster query building.

« I usually use several websites that I already know. » FG 4, Female, GP, 34 y

For example, residents mainly used French medical school sites. Beyond this short list, participants had a blurred vision of available information sources. Their knowledge of website characteristics was vague and confused, especially concerning health information databases.

Reliability was considered as relevant, but not fundamental criteria for choice. Their assessment of reliability was mostly based on intuition.

« We're smart enough, and have enough experience and training to make the difference. » FG 4, Male, GP, 52 y

As GPs, they felt incompetent or illegitimate for this task.

« Do we as GPs, have time to check if a website is trustworthy or not? Do we have the skills to check the validity of a website? » FG 5, Female, GP, 57 y

« Whether I can trust a paper or not? I don't have the expertise... » FG4, Female, GP, 45 y

However, several evaluation criteria were identified. People who had introduced them to the site, or who were involved in the site content, were taken into account. More credit was granted to institutions, like national health agencies, hospitals or universities.

« I've got more faith in a university or a hospital website. » FG 4, Male, GP, 58 y

The presentation and appearance of websites played a role in the participants' choices. The participants, and physicians more than residents, sought concordance with their existing knowledge.

« I refer to what I already know. If it's completely out of sync with what I already know, I tend to move on. » FG3, Male, GP, 56 y

This search for concordance involved both content and form, such as use of medical vocabulary.

« I think published papers are written in a certain way, which inspires trust. » FG4, Male, GP, 30 y

It rarely allowed for questioning of participants' knowledge and skills. The participants cited few objective criteria: an industry-free funding model, updating of data, assessment of the level of evidence, and referencing of data.

Participants stressed the importance of data presentation. They considered there was too much available data, and stressed the need for concise answers.

« I feel like I'm drowning in a sea of information. I'm afraid of not finding what I'm looking for. » FG 4, Male, GP, 30 y

They called for a data ranking system based on medical disciplines, age of data and relevance thereof. Participants regularly cited the English language and some websites charging for access as obstacles to seeking information online.

Participants sought information that was directly linked to their clinical practice. Through integration of an EBM care model, data seeking had to help the physicians deliver comprehensive care, into a patient-centred approach. Patient-centredness is a multifaceted concept involving several dimensions, such as clinical-patient relationship, patient as a unique person, patient information and patient empowerment [10].

« When I prescribe non common drugs to pregnant women, I systematically look on this website and I show them how to use it, so they can empower themselves. » FG 4, Female, GP, 34 y

« My decision will not be exclusively based on the information I may find online. The decision will be based on several things: this detailed information, the person in front of me and my relationship with this person. » FG 4, Male, GP, 52 y

Only a few of them were aware of the existence of general medicine guidelines databases, in French or foreign languages.

« A website for family physicians is missing. » FG3, Male, GP, 60 y

They were therefore ready to sacrifice some reliability on the altar of direct applicability of data.

« During the consultation, I'll accept less reliability as long as I find an answer which allows me to go forward. » FG5, Female, GP, 58 y

3.5. Information seeking process

The information seeking processes varied among participants, who emphasized the role of experience in building more valid queries.

« I perform the search my way. » FG4, Female, GP, 45 y

Most of their queries were built with natural language, without using controlled vocabulary. Participants were not aware of the role of key words and Boolean operators; especially among

physicians. They did not master the specific searching process of medical databases; and felt incompetent in their use.

« [Talking about Pubmed] It is really complicated, I find it hard to use [General approval]. » FG2, Female, Resident, 26 y

Google was the most frequently used search engine. Building their queries, participants integrated the type of websites to search for the information they needed, or the type of documents.

« Well, we all have our own methods (...): you enter the search term, and add "pdf". This accesses more relevant and scientific papers. » FG3, Male, GP, 60 y

Participants were fully aware of the impact query quality had on the validity of data. Nevertheless, querying was considered as a difficult process, and they were usually unsatisfied with their search results: useful answers, when they existed, were lost in a mass of non-relevant data. They felt they had circumvented the so-called official querying process, and felt guilty of not mastering available tools.

« I really don't feel like I'm doing it how I should. » FG1, Female, Resident, 29 y

3.6. Research context

Research process depended on time available.

« I don't use the same Websites and the same web search strategy when I have the time, for instance for teaching, or when I help a doctoral student with his/her thesis. In these cases, I adopt more structured pathways, i.e. using web search tools, and databases etc. » FG5, Male, GP, 43 y

Participants differentiated the fast and precise queries they performed during their working hours from the more global queries they could do at home. Patient involvement in the search process varied. Patients were usually informed of ongoing information research performed either in real time or otherwise.

« I would feel like I was betraying his/her trust, if I hid the fact I was looking up information. As if the consultation was for both of us. » FG5, Male, GP, 47 y

Patient involvement depended on the patient, the current health problem and the potential risk of acute anxiety.

« I feel that including them in the search gets them involved. » FG5, Female, Locum GP, 27 y

4. Discussion

4.1. Summary of main findings

Participants described a wide range of research topics, covering all general medicine core competencies, and especially patient-oriented topics. They used a limited list of websites. Participants were not confident about their ability to assess the quality of the information they found. Their assessment of data quality was based on intuition, and they mainly sought concordance with their existing knowledge. The way of presenting the data was considered very important. Participants were looking for information that was directly linked to their clinical practice. Information seeking processes varied among participants. They felt they had not mastered query building for conducting searches, and were aware the impact this shortcoming had on the quality of their search for information.

4.2. Comparison with existing literature

Several studies have already tried to build a taxonomy of questions asked by GPs' during their clinical exercise. They agree on the key part of diagnostic- and medication-related questions [7,11]. In the study conducted by Gonzalez-Gonzalez et al., more than half the questions related to diagnostic topics, and one in four to treatment topics. Our study shows that the range of questions asked by GPs' during their consultations is much wider. They reflect all general medicine core competencies, and especially a patient-centred approach [12]. Based on our 20 years experience of developing the CISMef catalogue [13], most of search engines and databases are disease-centred. The implementation of the patients' perspective in search engines and databases may help GPs' in patient care.

In our study, participants clearly pointed out the importance of evidence-based medicine. Their answers show that participants are infused with EBM culture. However, our findings demonstrate that a large gap remains between intention and practice. First, participants mostly sought concordance with their existing knowledge. They used clinical queries to test the validity of a hypothesis by seeking only confirming evidence. Positive hypothesis testing in the Internet is a well-known topic, especially among patients [14]. Among residents and GPs', the use of online resources might be strategically thought: they knew the overarching "answer" and were using the tools for the details. It could also be explained by the lack of skills to perform a literature search and the fear of information overload. Therefore, search engine developments may be needed to provide disconfirming evidence. Participants were aware of the role of validated data in the care of patients, but felt incompetent when seeking information, despite the existence of medical informatics training programmes including Internet search for health information [15]. The situation is multi-factorial. Many obstacles to query building and solution finding have already been identified in the literature. The amount of time necessary to find information, difficulties in reformulating the original question and finding an optimal search strategy, lack of a good source of information, uncertainty as to whether all relevant information has been found, and inadequate synthesis of any pieces of evidence into a clinically useful approach are the most cited [2,7]. In 2007, a web log analysis was undertaken in a meta-search engine covering 150 health resources and a variety of guidelines. It showed that most queries were built using a single search term and no Boolean operator [16]. The format of the requested data was also cited as a relevant topic. When asking GPs for factors which facilitated Internet use for information seeking in clinical practice, most asked for evidence-based summaries [4]. Using synthesized results of systematic evidence surveillance allowed physicians to answer more questions and change clinical decisions more often [17]. The best format for this synthesized information remains unclear [18]. In an EBM model care context, another obstacle for seeking information cited by participants was assessment of the reliability of data. Participants, and especially physicians, felt incompetent for this task. Their assessment was mainly intuitive, and focused on external validation, such as the authority of the institutional journal hosting the website. This latter is the major criterion of the Health On the Net code of conduct, the world reference to certify health information websites on the Internet [19]. Internal validation focused on the content and rigor of research. It remained incidental, despite its importance in the methods of evidence-based practice [20]. This feeling of incompetence was sometimes associated with a feeling of illegitimacy, and some participants experienced the impression that some websites, and especially medical databases, were not made for them. This clearly shows a gap between an idealized academic model of searching and real world practicalities [21].

4.3. Strengths and limits

This study has some limitations. Despite a large number of participants, only a few worked in single medical practices, and several were involved in residency teaching, which may affect the diversity of our sample. The responses of participants may have been affected by the fact that the two interviewers were involved in residency teaching. Nevertheless, their answers did not indicate that they were reluctant to report behaviors they considered themselves as inappropriate. Moreover, the researchers who participated in the data collection were not involved in the junior doctors' clinical assessments.

The diversity of participants, the facts that the analysis was conducted by several researchers and that the data saturation was achieved allowed us to ensure credibility. This study involved 35 residents and GPs, from different geographical areas and with various practice settings, which ensures good transferability.

4.4. Perspectives

Physicians and residents in general medicine are fully aware of the importance of evidence-based medicine and the need for objective and reliable information. Our study shows their difficulties in identifying this kind of data, and suggests ideas for improvement. Our findings reflect the recent changes affecting the profession of GPs, and especially the importance of knowledge access. Available tools have to change in order to fill the gap with real-world clinical practice, for example by integrating the patient's perspective. For example, this could involve a symptom or complaint focus in search engines or databases. Scientific information could also be systematically combined with figures, illustrations and summarized information in a format that patients understand. Information synthesis should be developed. We also need to precisely identify the information needs of GPs, who work in a comprehensive and complex context of multimorbidity. We need to raise GPs awareness of "research culture". Training interventions to improve information management have positive effects on GPs' behaviors, and need to be developed [22,23]. Continuous medical education to acquire proper information searching skills on the Internet seems mandatory.

Competing interests

The authors declare they have no competing interests.

Authors' contributions

MS, AM, and SJD had the initial idea of the study. MS, QF, and AM conceived the topic guide. MS, QF, and AM collected the data and performed analysis. MS drafted the manuscript. NG, GK, AM and SJD made substantial enhancement to it. All the authors approved the final manuscript.

Availability of supporting data

Supporting data are available in an open access repository (LabArchives). They include all the transcribed verbatim, in French. Data are available at the following URL: <https://mynotebook.labarchives.com/share/Cismef%2520/MjAuOHw2NDQ3MS8xNi9UcmVITm9kZS8yMTU1NzkwODQ4fDUyLjg=>.

Summary points

What was already known on the topic

- GPs' face a wide range of clinical situations during their practice. Nevertheless, despite the use of online information sources, some clinical questions remain unanswered.
- Time constraints, lack of skills to perform a literature search, information overload and heterogeneity in information quality are the most frequently cited barriers for the use of Internet for information seeking.

What this study added to our knowledge

- Information needs of GPs, who work in a comprehensive and complex context of multimorbidity, have to be precisely identified.
- Available tools have to change in order to fill the gap with real-world clinical practice, by integrating a patient-centred approach.

Acknowledgements

The authors would like to thank all the residents and GPs who took part at this study. The authors are grateful to Nikki Sabourin-Gibbs, Rouen University Hospital, for reviewing the manuscript in English.

References

- [1] D.L. Sackett, W.M. Rosenberg, J.A. Gray, R.B. Haynes, W.S. Richardson, Evidence based medicine: what it is and what it isn't, *BMJ* 312 (1996) 71–72.
- [2] H.C.H. Coumou, F.J. Meijman, How do primary care physicians seek answers to clinical questions? A literature review, *J. Med. Libr. Assoc.* 94 (2006) 55–60.
- [3] N.L. Bennett, L.L. Casebeer, R. Kristofco, B.C. Collins, Family physicians' information seeking behaviors: a survey comparison with other specialties, *BMC Med. Inform. Decis. Mak.* 5 (2005) 9.
- [4] E. Bernard, M. Arnould, O. Saint-Lary, D. Duhot, G. Hebbrecht, Internet use for information seeking in clinical practice: a cross-sectional survey among French general practitioners, *Int. J. Med. Inf.* 81 (2012) 493–499.
- [5] K. Ramos, R. Linscheid, S. Schafer, Real-time information-seeking behavior of residency physicians, *Fam. Med.* 35 (2003) 257–260.
- [6] K. Davies, J. Harrison, The information-seeking behaviour of doctors: a review of the evidence, *Health Inf. Libr. J.* 24 (2007) 78–94.
- [7] R. Krueger, M. Casey, *Focus Groups: A Practical Guide for Applied Research*, 4th edition, SAGE Publications, London, 2009.
- [8] E.W. Westberg, A.M. Randolph, The basis for using the internet to support the information needs of primary care, *J. Am. Med. Inform. Assoc.* 6 (1) (1999) 6–25.
- [9] J.W. Ely, J.A. Osheroff, M.H. Ebell, M.L. Chambliss, D.C. Vinson, J.J. Stevermer, E.A. Pifer, Obstacles to answering doctors' questions about patient care with evidence: qualitative study, *BMJ* 324 (2002) 710.
- [10] I. Scholl, J.M. Zill, M. Härter, J. Dirmaier, An integrative model model of patient-centeredness. A systematic review and concept analysis, *Plos One* 9 (9) (2014) e107828.
- [11] A.I. González-González, M. Dawes, J. Sánchez-Mateos, R. Riesgo- Fuertes, E. Escortell-Mayor, T. Sanz-Cuesta, T. Hernández-Fernández, Information needs and information-seeking behavior of primary care physicians, *Ann. Fam. Med.* 5 (2007) 345–352.
- [12] J. Allen, B. Gay, H. Crebolder, J. Heyrman, I. Svab, P. Ram, The European definitions of the key features of the discipline of general practice: the role of the GP and core competencies, *Br. J. Gen. Pract.* 52 (2002) 526–527.
- [13] S.J. Darmoni, J.P. Leroy, F. Baudic, M. Douyère, J. Piot, B. Thirion, CISMeF: a structured health resource guide, *Methods Inf. Med.* 39 (1) (2000) 30–35.
- [14] V.O. Kayhan, Seeking health information on the web: positive hypothesis testing, *Int. J. Med. Inf.* 82 (2013) 268–275.
- [15] F. Charles, Faculté de Médecine et Pharmacie – Informatique Médicale & Technologies de l'information et de la communication [<http://medecine-pharmacie.univ-rouen.fr/informatique-medicale-technologies-de-l-information-et-de-la-communication-16308.kjsp?RH=1381224035996>].
- [16] E. Meats, J. Brassey, C. Heneghan, P. Glasziou, Using the turning research into practice (TRIP) database: how do clinicians really search? *J. Med. Libr. Assoc.* 95 (2007) 156–163.
- [17] B.S. Alper, D.S. White, B. Ge, Physicians answer more clinical questions and change clinical decisions more often with synthesized evidence: a randomized trial in primary care, *Ann. Fam. Med.* 3 (2005) 507–513.
- [18] H.C. Barry, M.H. Ebell, A.F. Shaughnessy, D.C. Slawson, F. Nietzke, Family physicians' use of medical abstracts to guide decision making: style or substance, *J. Am. Board Fam. Pract.* 14 (2001) 437–442.
- [19] C. Boyer, O. Baujard, V. Baujard, S. Aurel, M. Selby, R.D. Appel, Health on the net automated database of health and medical information, *Int. J. Med. Inf.* 47 (1997) 27–29.
- [20] R.J. Cullen, In search of evidence: family practitioners' use of the Internet for clinical information, *J. Med. Libr. Assoc.* 90 (2002) 370–379.
- [21] P. Younger, Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature, *Health Inf. Libr. J.* 27 (2010) 2–10.
- [22] K.E. Schifferdecker, V.A. Reed, K. Homa, A training intervention to improve information management in primary care, *Fam. Med.* 40 (2008) 423–432.
- [23] M.P. Gagnon, P. Pluye, M. Desmartis, J. Car, C. Pagliari, M. Labrecque, P. Frémont, J. Gagnon, M. Njoya, F. Légaré, A systematic review of interventions promoting clinical information retrieval technology (CIRT) adoption by healthcare professionals, *Int. J. Med. Inf.* 79 (2010) 669–680.

2.5. Conclusion

Cette étude qualitative réalisée auprès d'internes de médecine générale et de médecins généralistes français met en évidence de nombreux freins à la recherche d'informations en santé en ligne. Parmi ces freins figure la langue anglaise, qui semble constituer un repoussoir pour les médecins installés mais également pour les médecins en formation. Cette barrière de la langue n'est pas spécifiquement française, elle figure parmi les obstacles principaux à la recherche d'information dans une étude menée auprès de 500 médecins européens (75). D'après les données que nous avons recueillies, il semble que ce frein concerne aussi bien l'information en elle-même que son support (base de données bibliographiques ou outil d'aide à la décision, par exemple).

Dans les prochaines parties de notre travail, nous présenterons la conception et l'évaluation de deux outils développés par le service d'informatique biomédicale du CHU de Rouen et destinés à faciliter la recherche d'information en ligne chez les chercheurs et professionnels de santé francophones ne maîtrisant pas la langue anglaise. Il s'agit des outils suivants :

- LiSSa (Littérature Scientifique en Santé, . Cette base de données bibliographiques francophone en santé a été créée en 2014 par un consortium pluridisciplinaire et à l'aide d'un financement de l'Agence Nationale de la Recherche (ANR).
- Le CRBM (Constructeur de Requêtes Bibliographiques Médicales, , qui permet de créer simplement des requêtes bibliographiques complexes en combinant des termes MeSH en langue française. Ces requêtes permettent d'interroger CISMEF, LiSSa et PubMed.

3. LiSSA : Littérature Scientifique en Santé

3.1. Les enjeux d'une base de données bibliographiques francophone dans le domaine de la santé

3.1.1. La prédominance de la langue anglaise

Grâce au Web, les professionnels de santé francophones peuvent plus facilement accéder à une très large quantité d'information en français. Les documents de qualité de sources institutionnelles ou universitaires y sont très nombreux. Le département de l'informatique et de l'information médicale du CHU de Rouen en recense plus de 116 000 via le catalogue et index des sites en langue française (CISMeF) (29). Ces dernières années, la langue anglaise a cependant encore accru sa prédominance en tant que langue véhiculaire des sciences de la santé, faisant presque disparaître les autres langues des bases de données. Ainsi, la proportion de publications en français référencées dans PubMed est passée de 12% dans les années 1960 à 1% actuellement (Figure 2).

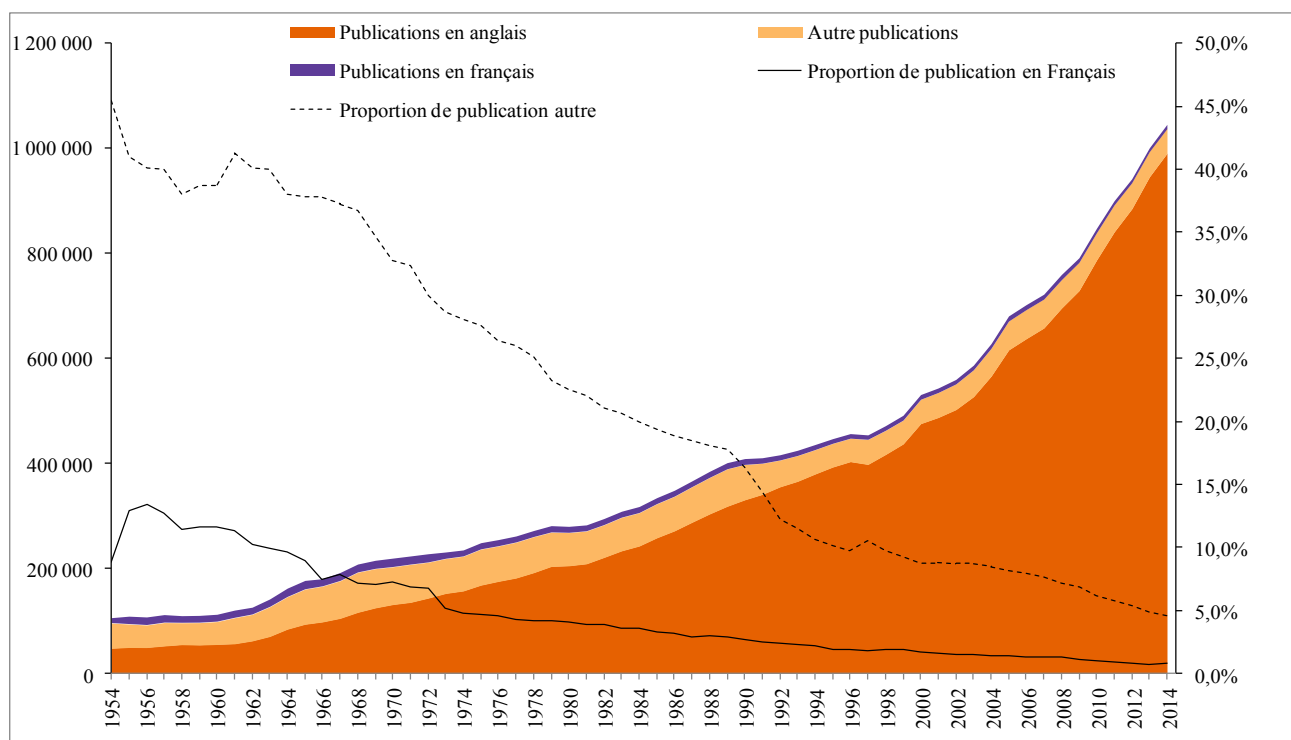


Figure 2 : nombre et proportion de publications en français dans PubMed, par années
(Source : PubMed)

3.1.2. La disparition progressive de la production francophone

Depuis quelques années, de nombreuses revues en français basculent vers l'anglais, comme les revues IRBM (*Innovation and Research in Biomedical Engineering*) (76) et la Revue Neurologique (77). Certaines conservent la revue en français mais créent une revue « sœur » en anglais, qui seule sera indexée dans PubMed, comme les Annales Françaises d'Anesthésie et de Réanimation (78). Concernant la Revue Neurologique, citons une partie de l'éditorial de JP. Azulay (77) qui résume assez bien la situation critique des revues scientifiques publiant en français : « 2016 verra le passage à l'anglais généralisé à tous les articles. C'est un tournant nécessaire que certains de nos lecteurs regrettent et qui me l'ont manifesté mais la seule revue de neurologie française indexée doit, pour survivre, développer son audience, progresser dans son contenu, être largement lue à l'étranger et particulièrement dans les pays anglophones afin que ses articles soient cités. Il est toutefois nécessaire que les lecteurs francophones puissent continuer à nous lire et que les auteurs qui ne maîtrisent pas l'anglais puissent continuer à soumettre des articles en français. (...) Avec cette stratégie nous espérons satisfaire tous nos lecteurs et en conquérir de nouveaux ».

Concernant le passage à l'anglais des revues françaises auparavant en français, un nouvel exemple très récent : les Archives de Pédiatrie, qui ont fait le même choix que la Revue

Neurologique. Citons Chabrol et al. (79) : « *Nous considérons que son rang actuel parmi les revues indexées de pédiatrie (117e/120) n'est pas un bon reflet de la qualité de la pédiatrie française et qu'une réforme en profondeur de notre revue devient nécessaire pour améliorer cette visibilité. Une réflexion a donc été menée au cours des derniers mois pour proposer une évolution des Archives de Pédiatrie qui réponde à ce double objectif : (a) optimiser la visibilité internationale et l'attractivité de notre revue pour les pédiatres non francophones, grâce à une augmentation de l'impact factor ; et (b) préserver son rôle pédagogique actuel dans les pays francophones.* ». Les publications en français disparaissant des bases de données bibliographiques, elles disparaissent donc des référencements bibliométriques. Publier en français est donc moins valorisé par les pouvoirs publics, notamment via l'indice bibliométrique SIGAPS (80) et moins valorisant pour les auteurs, aggravant encore le processus de décroissance.

3.1.3. Un niveau de maîtrise de la langue anglaise qui reste insuffisant

Malgré sa prédominance, l'anglais est un frein, voire un obstacle à la lecture pour de nombreux professionnels de santé français (81), et sans doute francophones. L'apprentissage de l'anglais médical existe dans la majorité des facultés de médecine en France, mais il est limité à quelques dizaines d'heures sur l'ensemble du cursus. Le niveau de compréhension de l'anglais des étudiants en médecine est améliorable, comme le montre l'étude de Lebrun et al. (82) effectuée sur cinq promotions successives d'étudiants en médecine de la faculté de médecine d'Angers. Durant cette période, le nombre d'étudiants de niveau B2 (niveau avancé défini par l'Union Européenne dans le cadre européen commun de références pour les langues) passait de 14 à 30 % lors du Test of English for International Communication (TOEIC). Actuellement, le niveau d'exigence dans les écoles d'ingénieurs est de 100 % pour le niveau B2, et l'éducation nationale souhaite même atteindre le niveau C1 (utilisateur expérimenté autonome) à la fin du Master 2. Les étudiants en médecine ont donc la nécessité de parfaire seuls cet apprentissage important pour leur formation initiale et continue. Certains verbatim issus de l'étude qualitative mentionnée dans le chapitre précédent reflètent cette difficulté : « *Bah, tous les sites en anglais, moi je ne regarde pas, je dis non...*(Interne de médecine générale) », « *Parfois, c'est en anglais, je pratique un petit peu, mais c'est plus difficile en consultation de lire en anglais et d'être rapide.* (Médecin généraliste) ».

3.1.4. Des outils existants qui ne répondent pas aux besoins des utilisateurs francophones

Des bases de données bibliographiques offrent depuis longtemps une large couverture de la littérature médicale. Aujourd'hui, PubMed constitue la principale porte d'accès gratuit vers la littérature scientifique en santé. De nouveaux acteurs sont également apparus sur le Web : des bases de données bibliographiques et bibliométriques scientifiques généralistes comme Scopus (URL : <http://www.scopus.com/>) (83) ou Web of Science (URL : <https://www.webofknowledge.com/>) (84), ou encore le moteur de recherche Google Scholar (URL : <https://scholar.google.com/>), qui présente des résultats intéressants en termes de couverture (85)(86). Ces outils sont toutefois développés en anglais : ils tiennent peu compte des spécificités des autres langues et ne disposent pas toujours d'interfaces adaptées aux utilisateurs non anglophones. En Amérique latine, la base de données bibliographique LILACS (Latin-America and Caribbean Health Sciences ; URL : <http://lilacs.bvsalud.org/en/>) (87) existe en trois langues : espagnol, portugais et anglais. En France, la principale base de données bibliographique en science, technologies et médecine, PASCAL, développée par l'Institut national de l'information scientifique et technique (INIST) depuis 1973 n'est plus mise à jour depuis fin 2014 (88). Il n'existe donc plus, à notre connaissance, de base de données bibliographique à jour, couvrant tous les domaines de la santé et prenant en compte les spécificités du français. Pour donner plus de visibilité, de vitalité et rendre accessible la littérature médicale et paramédicale en français, le projet BDBfr (base de données bibliographique en français), financé par l'Agence nationale de la recherche (ANR) dans le programme Technologies pour la Santé (TecSan), a visé à créer une base de données bibliographique en santé et en français, disponible gratuitement à travers un site internet « LiSSa, Littérature Scientifique en Santé » (URL : www.lissa.fr). Ce projet rassemble deux équipes académiques spécialisées en e-santé (le service d'informatique biomédicale du CHU de Rouen et le Centre d'Investigation Clinique Innovation Technologique de Lille) et trois acteurs industriels (un leader de l'édition scientifique - Elsevier-Masson - et deux sociétés privées françaises innovantes - Alicante et Sense-Gates -).

3.2. Présentation de la base de données LiSSa

3.2.1. Les données sources

Les données bibliographiques de LiSSa proviennent de deux sources : les éditeurs d'une part, et la base de données PubMed d'autre part.

Concernant PubMed, la *National Library of Medicine* (NLM) met gratuitement à disposition ses données, déposées dans le domaine public (NCBI). Pour chaque article, peu d'informations en langue française sont disponibles, le résumé et les mots-clés décrivant l'article étant exclusivement en anglais. Un programme spécifique a été développé afin de permettre une mise à jour quotidienne de LiSSa avec les données issues de PubMed.

Concernant les éditeurs, l'éditeur de journaux scientifiques proposant le plus grand nombre de titres en langue française est aujourd'hui Elsevier-Masson, l'un des partenaires du projet. Il couvre environ deux tiers des publications scientifiques en français. Certains des articles issus de ces revues sont indexés dans PubMed, les données fournies par l'éditeur permettent alors un enrichissement de leur description (affiliations détaillées de tous les auteurs, résumés en français). Des contacts avec plusieurs autres éditeurs ont également été pris, afin d'enrichir l'offre bibliographique de la base. Enfin, les médecins généralistes constituant une cible privilégiée de LiSSa, les articles issus de la revue exercer (la revue du Collège National des Généralistes Enseignants) sont également indexés dans la base (89).

A ce jour, plus de 600 000 références en français, publiées dans 2769 journaux, ont été extraites de PubMed et intégrées dans LiSSa. Elsevier-Masson a fourni les données de plus de 279 500 références publiées dans 125 revues du monde de la santé en langue française. En mars 2016, les 43 disciplines de l'Encyclopédie Médico-Chirurgicale (EMC), soit 7698 références, ont été intégrées dans la base. Au total, LiSSa compte 869 834 références d'articles (ou de chapitres d'ouvrage). L'intersection entre le corpus francophone de PubMed et les ressources d'Elsevier-Masson est de 157 568 références, soit 18% de l'ensemble des références de LiSSa.

3.2.2. Le modèle de données et le moteur de recherche

Le département d'informatique et d'information médicales du CHU de Rouen développe

depuis de nombreuses années un catalogue de ressources (29) et un moteur de recherche associé (4). Ces derniers ont constitué les éléments de base pour la création de LiSSa. Le modèle de données du catalogue CISMeF était initialement prévu pour décrire des ressources en ligne (90). Il a dû être modifié et enrichi de manière à pouvoir gérer des articles scientifiques et de formation continue contenus dans les deux principales sources de données initiales : l'ensemble des références bibliographiques en langue française de PubMed et de Elsevier-Masson. Nous nous sommes appuyés sur le Dublin Core (39) et le Learning Object Metadata (IEEE 1484.12.1-2002) à cette fin.

Le modèle de données sur lequel reposait le moteur de recherche sémantique du catalogue CISMeF ayant été modifié, ce dernier a été transformé en moteur générique capable non plus d'interroger une base de données intégrant uniquement des ressources en ligne, mais n'importe quelle base de données contenant tout type de documents, et en l'occurrence pour LiSSa des citations d'articles.

3.2.3. Les fonctionnalités de LiSSa

De nombreuses fonctionnalités ont été mises en place au sein de la base de données bibliographiques.

Le moteur de recherche utilise une base de connaissances pour identifier les termes médicaux. Par exemple une recherche sur « l'arythmie cardiaque » permet d'obtenir les références indexées avec le terme « troubles du rythme cardiaque ». De plus il permet l'explosion des termes, c'est-à-dire que la recherche concernera également les références indexées avec les « fils » du terme identifié. En pratique, si le moteur de recherche identifie dans la requête « troubles du rythme cardiaque », il permettra d'obtenir également les références indexées avec les termes « extrasystoles » ou « tachycardie ». Cette base de connaissances est constituée des terminologies de références mais également d'apports développés par l'équipe de documentalistes du SIBM comme les métatermes qui permettent d'identifier l'ensemble des références d'une même discipline médicale.

En termes d'interface, un outil de recherche avancée permet un usage facilité des opérateurs booléens.

Une formule spécifique de tri des références a été mise en place. Elle met en avant les revues de la littérature et les recommandations. Sa pertinence a été validée par des tests

utilisateurs. Il s'agit d'une formule complexe, utilisant de nombreux critères (titre, pondération MeSH, date, présence d'un résumé).

Des exports sont possibles vers des logiciels de gestion bibliographique, comme EndNote ou Zotero. Des exports sous forme de tableau ou de page web simplifiée ont également été mis en place, suite à des demandes d'utilisateurs.

Un utilisateur effectuant une recherche sur LiSSa peut, en un clic, découvrir les résultats de sa requête dans d'autres bases de données, comme CISMef, PubMed, Google Scholar ou LILACS.

Un outil de cartographie a été mis en place, il permet de localiser les institutions des auteurs des références.

Enfin, le score bibliométrique SCIMAGO a été intégré à la base. Il s'agit d'une mesure de l'influence scientifique des revues académiques tenant compte à la fois du nombre de citations obtenues par la revue (comme le fait l'Impact Factor), mais aussi de l'importance des journaux d'où proviennent ces citations (91).

L'interface de LiSSa sur ordinateur et téléphone mobile est visible sur la figure 3.

3.2.4. Formation des utilisateurs de LiSSa

Dès septembre 2015, LiSSa a été intégrée à la formation des futurs médecins de la faculté de médecine de Rouen, en deuxième année et en troisième cycle de médecine générale. Une action spécifique de formation a été menée dans les départements de médecine générale de Lille et Nice, ainsi qu'à la Cité de la Santé et dans une école de kinésithérapie de Berck. L'effort de formation dans les départements de médecine générale sera poursuivi dans les mois et années à venir. Une formation des enseignants de médecine générale de l'université de Lyon est prévue en septembre, tandis que les chefs de clinique de médecine générale de l'ensemble des universités françaises bénéficieront d'une formation aux outils développés par le SIBM (incluant LiSSa) lors d'un séminaire en novembre 2017. Enfin, des tutoriels vidéo ont été mis en ligne afin de faciliter l'utilisation de l'outil (<http://www.chu-rouen.fr/cismef/aide-pour-lissa/>).

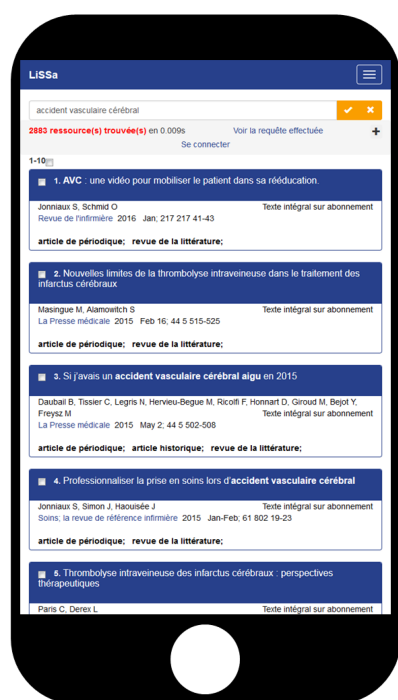
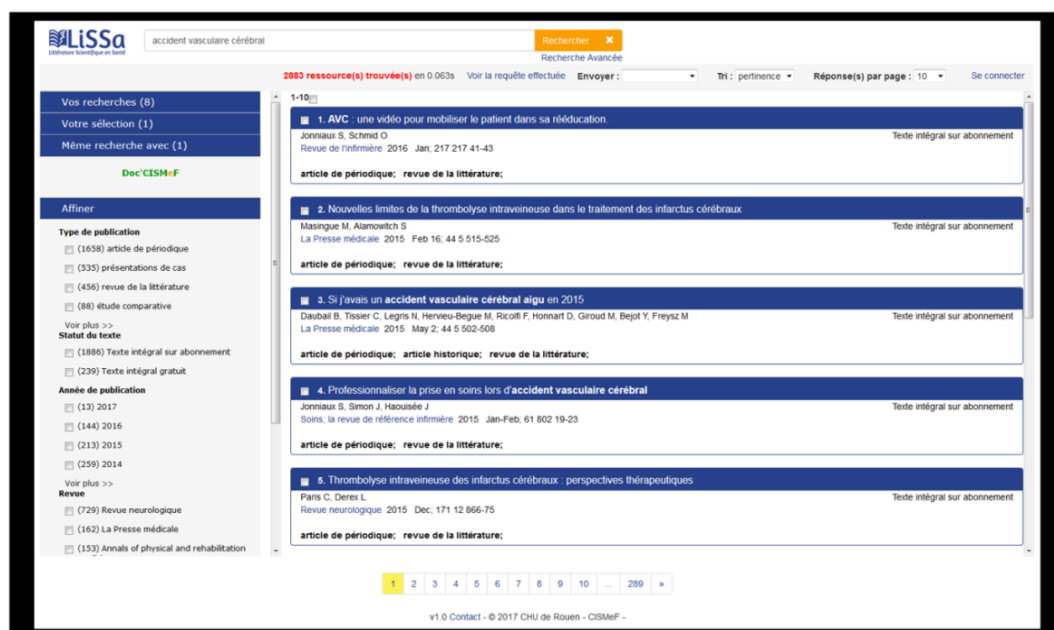


Figure 3 : Copies d'écran de l'interface de LiSSa pour ordinateur et téléphone mobile.

3.2.5. Utilisation et évaluation de l'outil

Environ 13 500 pages de LiSSa sont consultées chaque jour et plus de deux millions de pages ont été consultées entre janvier et juin 2017. En moyenne, 300 requêtes sont effectuées par jour, et on compte 440 utilisateurs uniques quotidiens.

Depuis la création de l'outil, plusieurs phases d'évaluation se sont déroulées, sous la direction du laboratoire d'ergonomie EVALAB, situé à Lille. Une première phase de tests utilisateurs préliminaires s'est déroulée en novembre 2014, lors du congrès du Collège National des Généralistes Enseignants de Lille, auprès de sept enseignants généralistes. Ces tests ont confirmé l'intérêt des participants pour une base de données bibliographiques en langue française. Les participants ont également exprimé leur confiance dans la validité scientifique des résultats de recherche présentés. La majorité a néanmoins estimé que des améliorations étaient nécessaires pour permettre un usage régulier de l'outil. Les principales difficultés rencontrées concernaient la compréhension de la recherche exécutée, l'affinement des résultats et l'accès aux résumés et aux articles. A la suite de cette évaluation, des recommandations ont été émises. Elles ont permis la mise en place de plusieurs modifications des fonctionnalités de l'outil. Une évaluation des formules de tri des références a été effectuée début 2017 auprès d'internes et de médecins généralistes de la région lilloise. Elle a permis la création de la formule de tri actuelle. Enfin, une évaluation ergonomique a été effectuée par les membres de l'équipe EVALAB, elle a permis d'élaborer 45 fiches problèmes et d'identifier 61 points d'amélioration.

3.2.6. Publications

Sont présentées ci-après trois publications relatives à la base de données bibliographiques LiSSa. La première, publiée dans le Journal of Medical Internet Research, se propose d'évaluer l'intérêt d'un moteur de recherche multilingue permettant d'interroger le sous-ensemble francophone des références de PubMed, notamment en termes de précision et de rappel. Les deux publications suivantes, issues de la Revue du Praticien et de la Presse Médicale, présentent la base de données LiSSa, les enjeux de sa création et ses fonctionnalités.

3.3. Un moteur de recherche permettant d'accéder aux sous-ensembles monolingues de PubMed. Preuve de concept et évaluation en français.

3.3.1. Contexte

La base de données PubMed contient de nombreuses références rédigées dans des langues différentes de l'anglais. Plusieurs outils permettent actuellement aux non-anglophones de requêter dans PubMed dans leur langue maternelle, voire de n'y sélectionner que les références rédigées dans leur langue maternelle. Cependant, les solutions existantes pour accéder à ces articles en utilisant la langue dans laquelle ils sont écrits demeurent peu convaincantes, notamment en termes de fonctionnalités disponibles.

3.3.2. Objectif

Proposer un moteur de recherche, appelé MultiLingual PubMed, qui permettra l'accès à un sous-ensemble de PubMed dans une langue, et évaluer sa qualité pour la version française (MultiLingual PubMed-French).

3.3.3. Méthode

Pour créer cet outil, des traductions du MeSH ont été enrichies (en y ajoutant des synonymes et des traductions en français) et intégrées au sein d'un portail terminologique. Toutes les références de PubMed en français ont été extraites de PubMed et incluses dans le moteur de recherche, en utilisant un parseur dédié. Des sous-ensembles de PubMed dans plusieurs autres langues européennes ont également été ajoutés à notre base de données. La précision et le rappel ont été mesurés sur les 20 premiers résultats de 20 requêtes sélectionnées aléatoirement parmi les requêtes les plus fréquentes effectuées sur CISMef. BabelMeSH, MultiLingual PubMed-French et 3 stratégies différentes de PubMed ont été comparées en recherchant la littérature rédigée en français. Les résultats ont été évalués selon la pertinence du titre et de l'abstract, l'évaluateur ne connaissant pas la stratégie de recherche utilisée.

3.3.4. Résultats

Plus de 650 000 citations de PubMed en français ont été intégrées dans l'outil. Les temps de réponse se situaient tous sous le seuil d'utilisabilité de deux secondes. Deux stratégies de recherche (MultiLingual PubMed-French et une des stratégies de PubMed) ont montré une grande précision (respectivement 0,93 et 0,97), mais le rappel était 4 fois supérieur pour le MultiLingual PubMed-French.

Une présentation détaillée de la méthode et des résultats de cette étude figure ci-après.

Original Paper

A Search Engine to Access PubMed Monolingual Subsets: Proof of Concept and Evaluation in French

Nicolas Griffon^{1,2,3,4}, MD PhD; Matthieu Schuers^{1,5}, MD; Lina Fatima Soualmia^{1,2}, PhD; Julien Grosjean¹, MSc; Gaétan Kerdelhué¹; Ivan Kergourlay¹, MSc; Badisse Dahamna¹, MSc; Stéfan Jacques Darmoni^{1,2}, MD PhD

¹CISMeF, TIBS, LITIS EA 4108, Rouen University Hospital, Normandy, Rouen, France

²U1142, LIMICS, Inserm, Paris, France

³UMR_S 1142, LIMICS, UPMC Univ Paris 06, Sorbonne Universités, Paris, France

⁴LIMICS (UMR_S 1142), Univ. Paris 13, Sorbonne Paris Cité, Villetaneuse, France

⁵Département de Médecine Générale, Université de Rouen, Rouen, France

Corresponding Author:

Nicolas Griffon, MD PhD

CISMeF

TIBS, LITIS EA 4108

Rouen University Hospital, Normandy

Unité d'informatique clinique - bâtiment Pillore

1 rue de Germont

Rouen, 76031

France

Phone: 33 232885726

Fax: 33 232888909

Email: nicolas.griffon@chu-rouen.fr

Abstract

Background: PubMed contains numerous articles in languages other than English. However, existing solutions to access these articles in the language in which they were written remain unconvincing.

Objective: The aim of this study was to propose a practical search engine, called Multilingual PubMed, which will permit access to a PubMed subset in 1 language and to evaluate the precision and coverage for the French version (Multilingual PubMed-French).

Methods: To create this tool, translations of MeSH were enriched (eg, adding synonyms and translations in French) and integrated into a terminology portal. PubMed subsets in several European languages were also added to our database using a dedicated parser. The response time for the generic semantic search engine was evaluated for simple queries. BabelMeSH, Multilingual PubMed-French, and 3 different PubMed strategies were compared by searching for literature in French. Precision and coverage were measured for 20 randomly selected queries. The results were evaluated as relevant to title and abstract, the evaluator being blind to search strategy.

Results: More than 650,000 PubMed citations in French were integrated into the Multilingual PubMed-French information system. The response times were all below the threshold defined for usability (2 seconds). Two search strategies (Multilingual PubMed-French and 1 PubMed strategy) showed high precision (0.93 and 0.97, respectively), but coverage was 4 times higher for Multilingual PubMed-French.

Conclusions: It is now possible to freely access biomedical literature using a practical search tool in French. This tool will be of particular interest for health professionals and other end users who do not read or query sufficiently in English. The information system is theoretically well suited to expand the approach to other European languages, such as German, Spanish, Norwegian, and Portuguese.

(*J Med Internet Res* 2014;16(11):e271) doi:[10.2196/jmir.3836](https://doi.org/10.2196/jmir.3836)

KEYWORDS

databases, bibliographic; French language; information storage and retrieval; PubMed; user-computer interface; search engine

Introduction

MEDLINE, created by the US National Library of Medicine (NLM), is the most used medical bibliographic database in the world. Currently (as of September 3, 2014), it contains 21,515,657 citations [1] from 5650 indexed journals from 81 countries around the world. Each MEDLINE record is indexed with the NLM's controlled vocabulary, Medical Subject Headings (MeSH) [2].

MEDLINE is the largest component of PubMed [3], the freely accessible online database of biomedical journal citations and abstracts. In addition to MEDLINE citations, PubMed also contains [4]:

1. Citations not yet indexed with MeSH and added to MEDLINE (in-process citations or when supplied electronically by the publisher);
2. Some "old MEDLINE" citations that have not yet been updated with the current vocabulary and converted to MEDLINE status;
3. Citations to some additional life science journals that submit full text to PubMedCentral and receive a qualitative review by the NLM; and
4. Citations to author manuscripts of articles published by NIH-funded researchers.

On the same date (September 3, 2014), PubMed contained 24,157,837 citations [5].

Language may be an obstacle to access PubMed and all the information it contains [6]. Moreover, there is still demand for native-language articles [7]. Of the 28 different ways to access PubMed listed by Lu [8], only 2 help non-native-English speakers to query PubMed/MEDLINE in their native language: BabelMeSH [9,10] and Patient, Intervention, Comparison, Outcome (PICO) Linguist [10]. We developed a French MeSH browser linked to PubMed [11]; more than 500 users consult it daily and it is taught in half of French medical schools. These 3 tools rely on MeSH translation in multiple languages to ease querying [12] and also rely on some metadata available in existing user languages to ease the results of browsing. Nevertheless, these 3 tools lack most PubMed functionalities (eg, advanced query builder, filters, citation sensor); in truth, they just lack functionalities.

A multilingual search engine to access the PubMed/MEDLINE subset in any non-English language (eg, French, German, Spanish, or Norwegian) with advanced functionalities would be of great interest for any user who is not comfortable in English. The goal of this paper was to present such a tool (Multilingual PubMed) and evaluate the potential increase in information retrieval task performance by using the French version.

Methods

Material

To create an effective multilingual search engine, the basic framework must first be clearly specified. Several tools were initially developed before the beginning of this project. Since

1995, the Catalogue and Index of Online Health Resources in French (CISMeF=Catalogue et Index des Sites Médicaux de langue Française), a quality-controlled health gateway [13], describes and indexes using the MeSH thesaurus the main institutional health Web resources (documents and sites) available in French. A search engine, Doc'CISMeF, was developed in 2000 to allow querying of this gateway. In 2007, CISMeF began using several terminologies/ontologies to allow easier and more accurate indexing. The CISMeF team was prompted to develop a multiterminology portal (French/English) [14], which interoperated with the search engine. In 2012, a transition was made to establish a cross-lingual portal named Health Terminology/Ontology Portal (HeTOP) available in 23 different languages [15,16].

Recently, Doc'CISMeF was significantly updated in 2 ways: (1) it has become a generic tool able to integrate any metadata model and its related data; therefore, this tool is able to describe and index Web resources as well as PubMed citations and (2) the search engine has now been developed into a reliable multilingual tool, which is capable of searching by using multiple terminologies/ontologies in several languages.

These 2 improvements in Doc'CISMeF allow the search for PubMed citation in any language. The final steps to transform Doc'CISMeF into the desired Multilingual Access to Each PubMed Subset for Each Language (Multilingual PubMed or, for French, Multilingual PubMed-French), designed for non-English speakers to access the literature in languages other than English, consist of (1) completing translation of MeSH descriptors and MeSH qualifiers and (2) translating the interface.

For the 5 basic languages of this study, the translation of MeSH descriptors is available in Unified Medical Language System [17]. Several institutions have helped us to translate the remaining MeSH qualifiers and interface terms (see Acknowledgments).

Nevertheless, developing a search engine requires good knowledge of the language used. As a French team, we are unable to assume that the performance for all other languages is equivalent to that of French language.

Two improvements related to the MeSH thesaurus in Doc'CISMeF can be used in Multilingual PubMed [18]:

1. Bilingual (French and English) predefined queries (n=389) were defined by the CISMeF medical librarians (eg, the CISMeF term "natremia" was automatically transformed to the MeSH predefined query "sodium/blood"). These predefined queries will have to be translated to be applied to the other European languages of this project.
2. Superconcepts (or metaterms; n=126) are medical (sub)specialties or biological science disciplines (eg, cardiology or bacteriology) selected by the CISMeF chief librarian. For each superconcept, 1 semantic link was created with 1 or more MeSH descriptor and qualifier. For example, the metaterm *psychiatry* is associated with the MeSH descriptors *psychiatry* and *psychiatric hospital* that belong to a completely different tree structure within the MeSH. Superconcepts have been created to optimize information retrieval and to overcome the relatively

restrictive nature of MeSH descriptors. These superconcepts also need to be translated, but the basic semantic links between superconcepts and MeSH descriptors are language-independent.

The list of predefined queries and metaterms is available at the HeTOP cross-lingual terminologies/ontologies portal [15].

There are also some other improvements that are based on all the available metadata. It is possible to perform faceted browsing [19]. Facets list citation characteristics, which can be used to refine results. It works based on metadata (eg, year, country, resource type). Default ranking in Doc'CISMeF is based on a relevance score, whereas in PubMed, the most recent articles are ranked first: last in-first out (LIFO). Doc'CISMeF interprets users' queries to identify meaningful words and medical descriptors matching health terminologies, such as MeSH. It then returns results corresponding to these words and descriptors. It uses several criteria to score and rank the results: query words present in the title, descriptors present in major topics, publication date, and origin of the descriptors as they may have been assigned by librarians or programmatically by computers for less important resources. If title words and major topics perfectly matched the user's query, they score 100% and are consequently ranked. Imperfect matches score less depending on how many words and descriptors are present (eg,

MeSH minor topic). For results of equal scores, publication date is the determinant criterion.

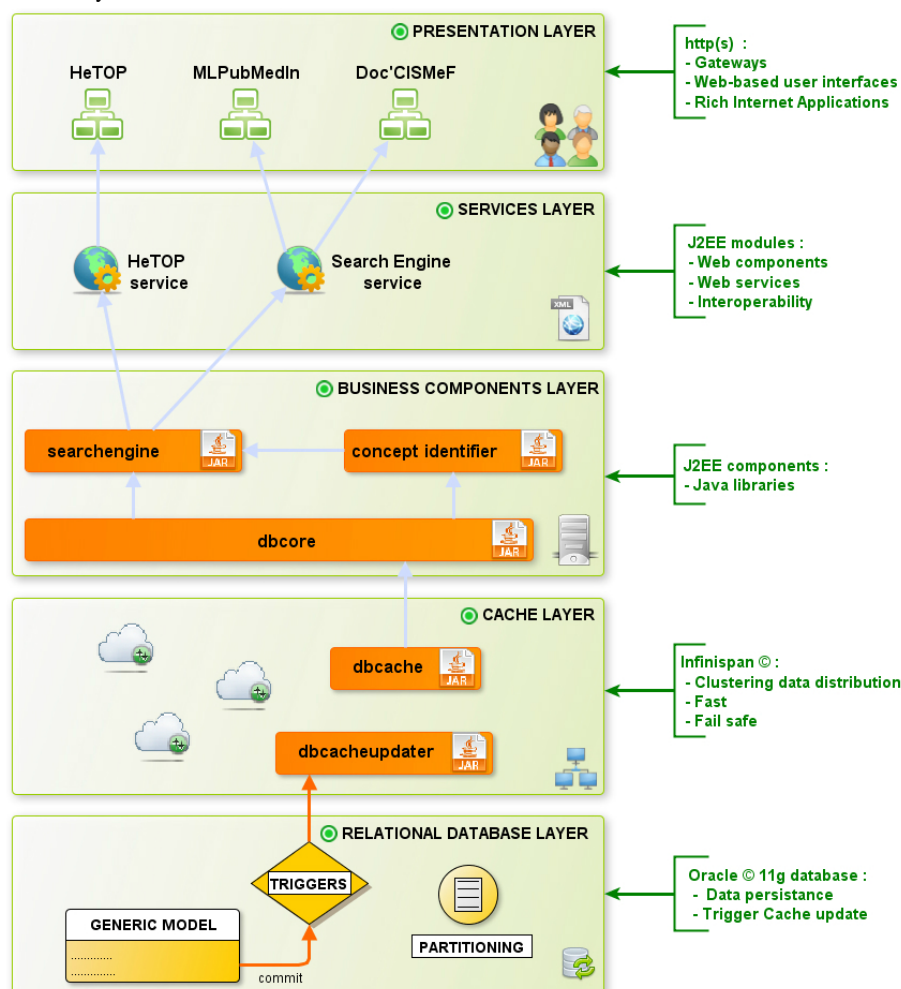
Some other improvements previously developed in Doc'CISMeF have not yet been implemented in the Multilingual PubMed-French search engine (see Discussion).

Architecture

The CISMeF generic information system is divided into 5 layers (Figure 1):

1. A relational database layer, which ensures data persistence and Structured Query Language query processing; execution plans are optimized by dynamic partitioning and local indexes.
2. A cache layer, which is devoted to clustering data distribution. It is based on random access memory and is distributed between several servers; therefore, it is very fast. Moreover, it avoids data loss after crashes.
3. A business components layer mainly composed of CISMeF Java libraries dealing with business logic.
4. A service layer hosting Web services that rely on business components and are, in turn, used by each client application of the information system.
5. A presentation layer that includes the gateways, Web-based user interfaces, rich Internet apps, and any client application of the service layer.

Figure 1. CISMeF information system.



Data Sources

The first and only version of Multilingual PubMed tested in this paper was the French version: the Multilingual PubMed-French search engine. Nevertheless, the whole framework should be able to manage any language as it does with French. To perform this study, all PubMed citations in French were extracted from PubMed and included in the Multilingual PubMed-French semantic search engine using a specific parser. Some PubMed citations were also extracted in other languages (ie, German, Spanish, Portuguese, and Norwegian) to ensure that our hypothesis of a multilingual framework was not unrealistic.

All MeSH descriptors and MeSH qualifiers were translated into French by the Technical and Scientific Information Department of the French National Institute of Health [20]. The CISMef team considerably enriched this translation with:

1. 25,501 synonyms and 689 ambiguous acronyms of MeSH descriptors;
2. 163 synonyms of MeSH qualifiers;
3. 20,887 translations of MeSH supplementary concepts (out of 209,326);
4. 27,295 synonyms of MeSH supplementary concepts;
5. 6037 translations of MeSH scope notes, which consist of MeSH descriptor definitions; and
6. 3918 “see also” relations among MeSH descriptors.

Evaluation

The 2 main limiting factors for such a tool are the response time, which must be low in order to be acceptable for end users, and the result quality (are there any results and are they relevant?). They were evaluated by measuring the response time and relevance of the first 20 results for 20 queries.

Queries were selected from the queries frequently run on Doc'CISMef (log-analysis). Twenty queries were randomly selected from those that retrieved results, were run more than 50 times in the previous 100 days, and did not involve advanced syntax.

Each query was run on Multilingual PubMed-French, BabelMeSH, and PubMed; the exact sentence from our log was pasted in each search engine. The corpus was limited to articles only in French language for each search engine. For PubMed, a second query was launched on the transliterated title (TT) field, hereafter referred to as the PubMed TT strategy. Finally, 2 PubMed algorithms for result sorting were tested: the classic LIFO and the recently proposed relevance ranking [21], which is probably more comparable to the Multilingual PubMed-French sorting algorithm. Thus, 5 strategies were tested: BabelMeSH, Multilingual PubMed-French, PubMed LIFO, PubMed relevance, and PubMed TT. Examples of queries are available in Table 1. PICO linguist was not evaluated because it works basically like Babel MeSH, but with a PICO interface (ie, with structured fields for patient, intervention, comparison, and outcome) which is designed for a specific query. The PubMed TT strategy was only studied with LIFO ranking because relevance ranking is based on the field in which the terms are found [21] and the PubMed TT strategy involved only 1 field: transliterated title. According to our data, PubMed LIFO and PubMed relevance ranking share 45% of their results. For PubMed TT strategy, this proportion rose to 88%.

The PubMed identifier (PMID) of the first 20 citations and the total number of citations were recorded by NG. For Multilingual PubMed-French, only response times were recorded. A medical librarian (GK) performed an assessment of citation relevance using a 3-modality Likert scale. He was blind to the search strategy that retrieved the resources to be evaluated. Therefore, it is possible to estimate the precision of each tool (ie, the proportion of relevant results in the results).

The relation between response time and number of citations was studied using the Pearson correlation coefficient. Two analyses were performed: for the strict one, only citations assessed as totally relevant were considered as true positives; for the relaxed one, both totally relevant and partially relevant citations were considered as true positives.

Table 1. Examples of queries in the different tools.

Search strategy	Example of query (for “constipation”) ^a
BabelMeSH ^b	constipation
Multilingual PubMed-French	constipation
PubMed ^c	constipation[All Fields] AND French[lang]
PubMed TT	constipation[TT] AND French[lang]

^a Shown is the query as written in the search box.

^b The checkbox “français” was checked.

^c Whatever the ranking used (relevance or LIFO), the query was the same.

Results

Technical Feasibility

Table 2 shows the number of PubMed citations for the main European languages (excluding English). As a proof of concept, PubMed citations of different languages were included in the CISMef information system: French (n=665,359), German

(n=7102), Portuguese (n=4497), Spanish (n=4297), and Norwegian (n=3764). The end user was able to choose 1 language and then perform a query in the same language. Thanks to our colleagues mentioned in the Acknowledgments, the Multilingual PubMed interface was translated into French, German, Portuguese, Spanish, and Norwegian. The same main metadata displayed by default in the PubMed database was similarly displayed in the Multilingual PubMed search engine,

as well as the indexing MeSH descriptors and qualifiers (Figure 2). Direct links to the full-text article in the same language were also presented to the end user (Figure 2), often via Digital Object Identifier [22]. The objective of creating a bibliographic database extracted from PubMed in several languages and available in 1 specific language was then completed as a proof of concept.

In the previous CISMeF gateway, the objective for response time was the following: all queries with 1 or 2 terms should be less than 2 seconds. Since 2001, this objective has been fulfilled for the 101,000 Web resources included in the CISMeF information system. The number of PubMed citations in French

versus CISMeF Web resources is approximately 6 times more. All response times for all queries were below the limit of 2 seconds (Table 3). As expected, there was a strong correlation between the response time and the number of citations retrieved ($r=.73$, 95% CI .42-.89). According to linear regression coefficient, this would lead to an unacceptable response time for queries that collect more than 20,000 citations. This evaluation of response time is an important step in determining the feasibility of this Multilingual PubMed-French search engine to create a bibliographic database in French for health professionals. The current version of Multilingual PubMed-French is available on the Internet [23].

Table 2. Number of PubMed citations for the main European languages.

Language	Number of PubMed citations
German	808,653
French	680,451 ^a
Italian	294,720
Spanish	302,287
Portuguese	85,839
Norwegian	35,712

^a The discrepancy between the number of French citations in PubMed and the number of French citations inserted in our information system comes primarily from some doubtful classifications in PubMed (ie, articles mislabeled as French).

Table 3. Response times for the 20 Multilingual PubMed-French queries.

Query (French/English)	Response time (seconds)	Number of PubMed citations
Allaitement maternel / breast feeding	0.42	633
Angine/pharyngitis	0.53	845
AVC/stroke	0.54	1704
BPCO/COPD	0.49	1503
Cigarette électronique / electronic cigarette	0.42	24
Constipation/constipation	0.27	531
Coqueluche / whooping cough	0.54	314
Gale/scabies	0.57	148
Hémochromatose/hemochromatosis	0.59	712
Hypertension/hypertension	1.03	8694
Lupus/lupus	0.34	1147
Maladie cœliaque / celiac disease	0.58	723
Maladie de Crohn / Crohn disease	0.69	1363
Nutrition/nutrition	1.13	8367
Psychiatrie/psychiatry	0.94	5602
Sarcoïdose/sarcoidosis	0.77	1721
Scoliose/scoliosis	1.10	669
Soins infirmiers / nursing care	0.86	5767
Tabac/tobacco	0.42	679
Toxoplasmose/toxoplasmosis	0.74	1238

Figure 2. Screenshot of the multilingual search engine.

The screenshot shows the MLPubMed(fr) search engine interface. At the top, there are navigation links: CISMef, À propos de, Sites et documents médicaux, Terminologies de Santé, Aide, and Se connecter. The search bar contains the query 'asthme'. Below the search bar, it indicates '3927 entrées trouvées en 2,31 s' and 'concept(s) identifié(s) : 0804493 - asthme asthme'. The left sidebar contains filters for 'Vos recherches (1)', 'Même recherche avec', 'Voir aussi', 'Votre sélection', 'Affiner', 'Type de publication' (with checkboxes for article de périodique, résumé en anglais, revue de la littérature, and présentations de cas), 'Statut du texte' (with checkboxes for Texte intégral sur abonnement and Texte intégral gratuit), and 'Année' (with checkboxes for 2007, 2006, 2005, and 2003). The main results area displays three search results, each with a title, authors, year, journal, and links to the full text. Annotations with arrows point to various features: 'PubMed data integration' points to the top right; 'History management' points to the left sidebar; 'Semantic relations' points to the left sidebar; 'Indexing' points to the MeSH terms for the second result; 'Realtime facet processing' points to the left sidebar; and 'Access to article content (DOI link)' points to the links at the bottom of the third result.

PubMed data integration

History management

Semantic relations

Indexing

Realtime facet processing

Access to article content (DOI link)

1. **Asthme de l'obèse un phénotype à part?**
Ben Kheder A, Maalej S 2014 pertinence 100%
Tunis Med. 2014 Jan;92(1);98
[*article de périodique:](#)
Voir l'indexation (12)
<http://www.latunisiemedicale.com/article-medicale-tunisie.php?article=2465>
<http://www.ncbi.nlm.nih.gov/pubmed/24879279>

2. **Atteinte des petites voies aériennes dans l'asthme.**
Charfi MR, Abouda M 2014 pertinence 100%
Tunis Med. 2014 Jan;92(1);101-3
[*article de périodique:](#) [*revue de la littérature:](#)
Voir l'indexation (9) MeSH: 'appareil respiratoire/anatomie pathologique
appareil respiratoire/physiopathologie
'asthme/anatomie pathologique
bronches/anatomie pathologique
bronches/physiologie
humains
muqueuse respiratoire/anatomie pathologique
muqueuse respiratoire/physiologie
phénotype
<http://www.latunisiemedicale.com/article-medicale-tunisie.php?article=2467>
<http://www.ncbi.nlm.nih.gov/pubmed/24879281>

3. **Rôle de l'épithélium bronchique dans la pathogénie de l'asthme.**
Rekik WK 2014 pertinence 100%
Tunis Med. 2014 Jan;92(1);99-100
[*article de périodique:](#) [*revue de la littérature:](#)
Voir l'indexation (6)
<http://www.latunisiemedicale.com/article-medicale-tunisie.php?article=2466>
<http://www.ncbi.nlm.nih.gov/pubmed/24879280>

1 2 3 4 5 6 7 8 9 10 ... 197 ►

v3.0 [Contact](#) - © 2014 CHU de Rouen - CISMef - Powered by Vaadin ►

Performance Evaluation

Table 4 displays the coverage of each query for each strategy. The PubMed TT strategy retrieved far fewer citations than the other 4 strategies with a total of 10,716 hits. Babel MeSH, Multilingual PubMed-French, and regular PubMed queries retrieved 50,894, 42,384, and 34,047 citations, respectively. Babel MeSH and Multilingual PubMed-French retrieved significantly more citations than relevance/LIFO PubMed (both $P=.03$; Mann-Whitney test).

The precision measured on the first 20 citations retrieved are listed (1) considering the citations assessed as highly relevant—the strict analysis—and (2) considering both the highly and the slightly relevant citations, the relaxed analysis (Table 5). The results were similar with 2 strategies that outperformed the other 3 in terms of precision: Multilingual PubMed-French and PubMed TT. These 2 strategies reached more than 90% precision with a slight advantage for PubMed TT ($P=.002$, Fisher test), whereas the other 3 barely reached 80% in the relaxed analysis.

Table 4. Coverage of queries according to the search engine.

Query (English)	Strategy (n)			
	BabelMeSH	Multilingual PubMed-French	PubMed ^a	PubMed TT
Breast feeding	639	633	3	50
Pharyngitis	2043	845	2326	82
Stroke	2671	1704	202	52
COPD	687	1503	25	275
Electronic cigarette	355	24	0	0
Constipation	574	531	666	144
Whooping cough	227	314	2	166
Scabies	136	148	4	102
Hemochromatosis	537	712	756	129
Hypertension	10,348	8694	12,580	1533
Lupus	2232	1147	2694	1150
Celiac disease	649	723	2	305
Crohn disease	1190	1363	2	758
Nutrition	4969	8367	6480	819
Psychiatry	4453	5602	5394	1940
Sarcoidosis	1522	1721	13	878
Scoliosis	651	669	3	159
Nursing care	14,360	5767	2867	888
Tobacco	1607	679	18	585
Toxoplasmosis	1044	1238	10	701
Total	50,894	42,384	34,047	10,716

^a Whatever the ranking used (relevance or LIFO), the number of citations retrieved was the same.

Table 5. Precision measured on the first 20 citations for each query for each strategy (analysis-relaxed analysis).

Query (English)	Strategy (precision)				
	Babel MeSH	Multilingual PubMed-French	PubMed LIFO	PubMed Relevance	PubMed TT
Breast feeding	0.55-0.75	1.00-1.00	1.00-1.00 ^a	1.00-1.00 ^a	1.00-1.00
Pharyngitis	0.00-0.00	0.65-0.70	0.00-0.00	0.00-0.00	0.50-0.60
Stroke	0.45-0.80	0.95-0.95	0.15-0.15	0.35-0.35	1.00-1.00
COPD	0.40-0.55	1.00-1.00	0.75-0.85	1.00-1.00	1.00-1.00
Electronic cigarette	0.10-0.10	0.35-0.40	— ^b	— ^b	— ^b
Constipation	0.45-1.00	1.00-1.00	0.50-0.95	0.75-1.00	1.00-1.00
Whooping cough	0.80-1.00	1.00-1.00	1.00-1.00 ^c	1.00-1.00 ^c	1.00-1.00
Scabies	0.90-1.00	1.00-1.00	0.25-0.25 ^d	0.25-0.25 ^d	1.00-1.00
Hemochromatosis	0.65-0.95	1.00-1.00	0.70-1.00	0.85-1.00	1.00-1.00
Hypertension	0.45-0.75	0.95-0.95	0.25-0.75	1.00-1.00	1.00-1.00
Lupus	0.70-1.00	1.00-1.00	0.75-0.95	1.00-1.00	1.00-1.00
Celiac disease	0.70-0.95	1.00-1.00	1.00-1.00 ^c	1.00-1.00 ^c	1.00-1.00
Crohn disease	0.80-0.90	1.00-1.00	0.50-1.00 ^c	0.50-1.00 ^c	1.00-1.00
Nutrition	0.45-0.70	0.70-1.00	0.25-0.60	1.00-1.00	1.00-1.00
Psychiatry	0.75-0.80	1.00-1.00	0.85-0.95	0.90-0.95	1.00-1.00
Sarcoidosis	0.70-0.90	1.00-1.00	0.92-0.92 ^e	0.92-0.92 ^e	1.00-1.00
Scoliosis	0.65-0.90	1.00-1.00	0.67-0.67 ^a	0.67-0.67 ^a	1.00-1.00
Nursing care	0.85-0.90	1.00-1.00	0.70-0.80	0.85-0.90	1.00-1.00
Tobacco	0.60-0.90	1.00-1.00	0.83-0.89 ^f	0.83-0.89 ^f	1.00-1.00
Toxoplasmosis	0.65-0.90	1.00-1.00	1.00-1.00 ^g	1.00-1.00 ^g	1.00-1.00
Total	0.58-0.79	0.93-0.95	0.57-0.74	0.79-0.83	0.97-0.98

^a Only 3 citations retrieved.^b No citation retrieved.^c Only 2 citations retrieved.^d Only 4 citations retrieved.^e Only 13 citations retrieved.^f Only 18 citations retrieved.^g Only 10 citations retrieved.

Discussion

Principal Results

The Multilingual PubMed-French strategy allows the retrieval of French PubMed citations with high precision and high coverage. Despite more than 600,000 PubMed citations that were introduced into the CISMef information system, the average server-side response time stayed below the threshold time (ie, 2 seconds), which is acceptable, in our experience, for the average health professional. The Multilingual PubMed-French search engine is now freely available on the Internet [23].

Babel MeSH allows users to access PubMed database in French. It provides many results for low precision and is, therefore, of

poor interest for French speakers. Moreover, the functioning of the search engine is not detailed anywhere; it is only possible to make rough free-term queries that are translated automatically by Babel MeSH. The Babel MeSH interface is limited; it is not possible to perform many queries in a short period of time.

The classic use of PubMed is not satisfying for French speakers, which is not surprising because it is supposed to work in English, not in French. The relatively high mean coverage masked enormous heterogeneity (from 0 to more than 12,000 citations). Precision was lower than that for Multilingual PubMed-French or PubMed TT. However, the relevance ranking provided consistently better precision for the first 20 results than LIFO ranking. There was no evidence to maintain LIFO ranking as default.

Using PubMed TT provides more precise results than using Multilingual PubMed-French. PubMed TT works like a free-text search engine that looks only in the title of the citation. Therefore, it is quite logical that precision was high. Nevertheless, PubMed TT recall must be lower than Multilingual PubMed-French recall based on the huge differences in coverage and slight differences in precision. Moreover, Multilingual PubMed-French and Babel MeSH constantly provided more than 20 results, whereas PubMed did not; some queries provided few or even no results. The poor coverage of PubMed TT is probably the consequence of difficulties managing French particularities (eg, elision, apostrophe, accentuation).

Limitations

This study has several limitations. First, this feasibility study only included server-side response time. The overall response time also depended on the end user's computer type and type of browser. Second, the MeSH thesaurus is incompletely translated. MeSH descriptors and qualifiers are frequently translated, but this is not the case for MeSH supplementary concepts. Therefore, querying at a MeSH supplementary concept level of accuracy is currently not possible in one's own language. It would require a huge amount of work to translate all MeSH supplementary concepts, even using natural language processing tools [24]. Last, English is by far the most frequently used language of scientific medical publications and publications in other languages are declining comparatively. Nevertheless, the latter are important because they can be more easily adapted to national contexts and be more understandable to a wider local audience. An automatic intelligible translation of English articles would definitely resolve this issue, but it is far beyond current technology [25].

For the “engine/pharyngitis” query, an unanticipated issue arose: PubMed (LIFO and relevance strategies) did not find any results and automatically translated the query to “angina” which has a different meaning. This affected the precision of these strategies by biasing it toward zero. Three sensitivity analyses were performed to evaluate bias importance: (1) the “engine” query was excluded from the analysis, (2) the “angine” query was considered as providing zero results for PubMed LIFO and PubMed relevance strategies, and (3) the 20 citations returned by this query for PubMed LIFO/relevance strategies were considered as fully relevant. In each scenario, the precision was significantly better for Multilingual PubMed-French and PubMed TT compared with PubMed LIFO/relevance strategies. Therefore, it is possible to conclude a very limited bias.

Perspectives and Future Challenges

One of the strengths of Multilingual PubMed lies in the multilingual CISMef information system. Because the MeSH thesaurus is already translated into multiple languages, it will be easy to make Multilingual PubMed for these languages. Nevertheless, the CISMef team is not familiar with the linguistic treatments that are necessary to interpret user queries in other languages than French; therefore, it is unlikely that the other

versions of Multilingual PubMed will achieve results comparable to those observed here. We are currently working with several international teams to extend Multilingual PubMed to German, Spanish, Portuguese, and Norwegian. According to the number of PubMed citations (see Table 1) for the main European languages, this feasibility study could be extrapolated as positive for other languages with fewer PubMed citations than French (ie, Spanish, Italian, Portuguese, and Norwegian) and one-third larger for German.

Several other improvements related to MeSH indexing in the search engine have not yet been implemented for PubMed citations; in particular, affiliation of MeSH qualifiers to MeSH supplementary concepts, indexing with MeSH concepts, and affiliation of MeSH qualifiers to MeSH concepts [26]. It will not be feasible to perform these tasks manually due to the size of the PubMed citations corpus in French (n=665,359). Therefore, the automatic indexing tool used in the CISMef quality-controlled health gateway will need to be enhanced to be applied to the Multilingual PubMed-French search engine.

As stated in the Introduction, it is possible to manage multiple kinds of resources in the CISMef information system. End users are now able to ask for mixed results: Web resources from CISMef and PubMed citations from Multilingual PubMed. Also, an interesting avenue for future development of Multilingual PubMed is to integrate bibliographic data from other sources. A partnership with Elsevier-Masson and 2 small French companies has now been launched. Elsevier will provide additional metadata for MEDLINE/PubMed citations (abstracts in French in particular), which currently are not yet available in the MEDLINE/PubMed database. Furthermore, Elsevier-Masson will provide metadata on journals not included in PubMed, but are included in other bibliographic/bibliometric databases (eg, Web of Science, EMBASE, and BIOSIS).

In addition to being more efficient than its competitors are, Multilingual PubMed is, in our opinion, more user friendly and provides more features. Nevertheless, there are still many interesting functionalities in PubMed (ie, related citations, history management) or in other tools [8] (graphical representation) that are not implemented in Multilingual PubMed. Future work may address these issues.

Conclusion

This feasibility study was carried out to create a multilingual search engine to query monolingual PubMed subsets was successful for French language and will be extended to the other main European languages. Bibliographical references from non-English publications can now be searched in a native language-friendly interface using multilingual data from the PubMed database and MeSH vocabulary.

Overall, this Multilingual PubMed tool could be of valuable interest for non-English-speaking health professionals who are unable to access PubMed.

Acknowledgments

The authors would like to thank Pablo Iriarte, Librarian at the University Hospital CHUV in Lausanne, Switzerland, for his decisive impulse for starting the project. The authors are grateful to Nikki Sabourin-Gibbs, Rouen University Hospital, for reviewing the manuscript in English. UMIT, Austria, Centre Tudor, Luxembourg, Hospital Italiano of Buenos Aires, Argentina, and the Norwegian Electronic Health Library graciously translated some MeSH terms and some interface terms into German, German and Portuguese, Spanish and Italian, and Norwegian, respectively.

Authors' Contributions

GK, LFS, and SJD had the idea of Multilingual PubMed and pushed for its development. JG, IK, and BD developed the whole application and optimized the architecture of the system. NG, GK, MS, and SJD conceived the evaluation. NG collected the data and performed analysis. GK made citation relevance assessment. NG and SJD drafted the manuscript. LFS and MS made substantial enhancement to it. All the authors approved the final manuscript.

Conflicts of Interest

None declared.

References

1. MEDLINE. Number of resources in MEDLINE URL: [http://www.ncbi.nlm.nih.gov/pubmed?cmd=PureSearch&db=pubmed&term=medline\[sb\]](http://www.ncbi.nlm.nih.gov/pubmed?cmd=PureSearch&db=pubmed&term=medline[sb]) [WebCite Cache ID 6S7HdKcK3]
2. Lipscomb CE. Medical Subject Headings (MeSH). Bull Med Libr Assoc 2000 Jul;88(3):265-266 [FREE Full text] [Medline: 10928714]
3. PubMed. URL: <http://www.ncbi.nlm.nih.gov/pubmed/> [accessed 2014-08-26] [WebCite Cache ID 6S7LTaVwv]
4. US National Library of Medicine. 2014. Fact sheet: MEDLINE, PubMed, and PMC (PubMed Central): How are they different? URL: http://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html [accessed 2014-08-26] [WebCite Cache ID 6S7Moo8Q8]
5. PubMed. Number of resources in PubMed URL: [http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=PureSearch&db=pubmed&term=all\[sb\]](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=PureSearch&db=pubmed&term=all[sb]) [WebCite Cache ID 6S7Howfws]
6. Sheets L, Gavino A, Callaghan F, Fontelo P. Do language fluency and other socioeconomic factors influence the use of PubMed and MedlinePlus? Appl Clin Inform 2013;4(2):170-184 [FREE Full text] [doi: 10.4338/ACI-2013-01-RA-0006] [Medline: 23874356]
7. Chneiweiss H. [Why a scientific review in French?]. Med Sci (Paris) 2014 Jan;30(1):7-8. [doi: 10.1051/medsci/20143001001] [Medline: 24472447]
8. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford) 2011;2011:baq036 [FREE Full text] [doi: 10.1093/database/baq036] [Medline: 21245076]
9. Liu F, Ackerman M, Fontelo P. BabelMeSH: development of a cross-language tool for MEDLINE/PubMed. AMIA Annu Symp Proc 2006:1012 [FREE Full text] [Medline: 17238631]
10. Fontelo P, Liu F, Leon S, Anne A, Ackerman M. PICO Linguist and BabelMeSH: development and partial evaluation of evidence-based multilanguage search tools for MEDLINE/PubMed. Stud Health Technol Inform 2007;129(Pt 1):817-821. [Medline: 17911830]
11. Thirion B, Pereira S, Névél A, Dahamna B, Darmoni S. French MeSH Browser: a cross-language tool to access MEDLINE/PubMed. AMIA Annu Symp Proc 2007:1132. [Medline: 18694229]
12. Nelson SJ, Schopen M, Savage AG, Schulman JL, Arluk N. The MeSH translation maintenance system: structure, interface design, and implementation. Stud Health Technol Inform 2004;107(Pt 1):67-69. [Medline: 15360776]
13. Darmoni SJ, Leroy JP, Baudic F, Douyère M, Piot J, Thirion B. CISMef: a structured health resource guide. Methods Inf Med 2000 Mar;39(1):30-35. [Medline: 10786067]
14. Portail Terminologique de Santé. URL: <http://pts.chu-rouen.fr/old/> [accessed 2014-08-26] [WebCite Cache ID 6S7GWqPTy]
15. Grosjean J, Merabti T, Griffon N, Dahamna B, Darmoni SJ. Teaching medicine with a terminology/ontology portal. Stud Health Technol Inform 2012;180:949-953. [Medline: 22874333]
16. Health Terminology/Ontology Portal. URL: <http://www.hetop.fr/hetop/> [accessed 2014-08-26] [WebCite Cache ID 6S7GyugRz]
17. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc 1998;5(1):1-11 [FREE Full text] [Medline: 9452981]
18. Douyère M, Soualmia LF, Névél A, Rogozan A, Dahamna B, Leroy JP, et al. Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. Health Info Libr J 2004 Dec;21(4):253-261. [doi: 10.1111/j.1471-1842.2004.00526.x] [Medline: 15606883]
19. Tunkelang D. Faceted search. Synthesis Lectures on Information Concepts, Retrieval, and Services 2009 Jan;1(1):1-80. [doi: 10.2200/S00190ED1V01Y200904ICR005]

20. Le MeSH bilingue anglais - français. URL: <http://mesh.inserm.fr/mesh/> [accessed 2014-08-26] [[WebCite Cache ID 6S7MtSGFa](#)]
21. Canese K. PubMed relevance sort. NLM Tech Bull 2013 Oct;394:e2.
22. Digital Object Identifier System. URL: <http://www.doi.org/> [accessed 2014-08-26] [[WebCite Cache ID 6S7MzU2OQ](#)]
23. MLEPubMedFr. URL: <http://doccismef.chu-rouen.fr/dc/?env=pubmed> [accessed 2014-08-26] [[WebCite Cache ID 6S7HHSAT](#)]
24. Merabti T, Soualmia LF, Grosjean J, Joubert M, Darmoni SJ. Aligning biomedical terminologies in French: towards semantic interoperability in medical applications. In: Mordechai S, Sahu R, editors. Medical Informatics. Croatia: InTech; Mar 2012:41-68.
25. Liu F, Fontelo P. Automatic translations of MEDLINE abstracts using Google Translate. In: AMIA Annu Symp Proc. 2010 Presented at: AMIA Annu Symp; Nov 13-17, 2010; Washington, DC p. 1148.
26. Darmoni SJ, Soualmia LF, Letord C, Jaulent MC, Griffon N, Thirion B, et al. Improving information retrieval using Medical Subject Headings Concepts: a test case on rare and chronic diseases. J Med Libr Assoc 2012 Jul;100(3):176-183 [[FREE Full text](#)] [doi: [10.3163/1536-5050.100.3.007](https://doi.org/10.3163/1536-5050.100.3.007)] [Medline: [22879806](#)]

Abbreviations

CISMeF: Catalogue et Index des Sites Médicaux de langue Française

LIFO: last in–first out

MeSH: Medical Subject Headings

NLM: National Library of Medicine

PMID: PubMed identifier

TT: transliterated title

Edited by G Eysenbach; submitted 03.09.14; peer-reviewed by H Chneiweiss, Z Lu; comments to author 01.10.14; revised version received 15.10.14; accepted 19.10.14; published 23.11.14

Please cite as:

Griffon N, Schuurs M, Soualmia LF, Grosjean J, Kerdelhué G, Kergourlay I, Dahamna B, Darmoni SJ

A Search Engine to Access PubMed Monolingual Subsets: Proof of Concept and Evaluation in French

J Med Internet Res 2014;16(11):e271

URL: <http://www.jmir.org/2014/11/e271/>

doi: [10.2196/jmir.3836](https://doi.org/10.2196/jmir.3836)

PMID:

©Nicolas Griffon, Matthieu Schuurs, Lina Fatima Soualmia, Julien Grosjean, Gaétan Kerdelhué, Ivan Kergourlay, Badisse Dahamna, Stéfan Jacques Darmoni. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.11.2014. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



EN DÉBAT

Les articles scientifiques écrits en français sont de plus en plus éclipsés par ceux rédigés en anglais, notamment dans les bases de données bibliographiques internationales. Ils constituent des ressources potentiellement utiles pour beaucoup de professionnels de santé. Un nouvel outil déjà riche de plus de 800 000 références vise à agréger l'ensemble de la littérature médicale en français.

Littérature scientifique en santé (LiSSa) : une base de données bibliographiques en français

Gâce au Web, les professionnels de santé francophones peuvent plus facilement accéder à une très large quantité d'information en français. Les documents de qualité de sources institutionnelles ou universitaires y sont très nombreux. Le service d'informatique biomédicale du centre hospitalier universitaire (CHU) de Rouen développe d'ailleurs depuis 1995 le Catalogue et index des sites en langue française (CISMeF)¹ qui en recense plus de 110 000.

Mais qu'en est-il des articles des revues scientifiques francophones ? C'est là que le bât blesse. Ces dernières années, l'anglais a encore accru sa prédominance en tant que langue véhiculaire des sciences de la santé, faisant presque disparaître les autres langues des bases de données (fig. 1). De nombreuses revues en français basculent vers l'anglais, comme la revue *IRBM*, consacrée aux technologies biomédicales,² ou conservent la revue en français mais créent une revue « sœur » en anglais, qui seule est indexée dans PubMed, comme les *Annales françaises d'anesthésie et de réanimation*.³ Les publications en français disparaissent des bases de données bibliographiques, et donc il en est de même des référencements bibliométriques. Publier en français est donc moins valorisé par les pouvoirs publics et moins valorisant pour les auteurs, aggravant encore le processus de décroissance.

Malgré sa prédominance, l'anglais est un frein, voire un obstacle à la lecture pour de nombreux professionnels de santé⁴ français, et sans doute francophones. Les publications en français restent le chemin le plus court pour l'acquisition des connaissances. De plus, les ressources institutionnelles ou pédagogiques ne peuvent pas résoudre tous les problèmes, soit qu'elles

NICOLAS GRIFFON*^{**}
NICOLAS.GRIFFON
@CHU-ROUEN.FR

MATTHIEU
SCHUERS*^{***}
XXXX

GAËTAN KERDELHUÉ*
XXXX

JULIEN GROSJEAN*
XXXX

STÉFAN JACQUES
DARMONI*^{**}
XXXX

* Service
d'informatique
biomédicale, CHU de
Rouen, Rouen, France

** Inserm U1142,
Paris ; Sorbonne
Universités, UPMC
université Paris-6,
Paris ; université
Paris-13, Sorbonne
Paris Cité,
Villetaneuse, France

*** Département de
médecine générale,
université de Rouen,
Rouen, France

n'atteignent pas un niveau de spécialisation suffisant, soit qu'elles ne sont pas adaptées aux situations auxquelles sont confrontés les professionnels. Enfin, les articles de synthèse restent plébiscités par le lectorat francophone.⁵ Les articles scientifiques en français ont donc encore toute leur utilité, et leur disparition serait potentiellement préjudiciable pour la mise à jour des connaissances et donc, *in fine*, pour la qualité des soins.

Des bases de données bibliographiques offrent depuis longtemps une large couverture de la littérature médicale. MedLARS a été créé en 1964,⁶ puis s'est développé sous la forme de Medline en profitant de l'essor des télécommunications. Aujourd'hui, PubMed est la principale porte d'accès gratuite vers la littérature scientifique en santé. De nouveaux acteurs sont ensuite apparus sur le web : des bases de données bibliographiques et bibliométriques scientifiques généralistes comme Scopus (<http://www.scopus.com/>),⁷ *Web of Science* (<https://www.webofknowledge.com/>),⁸ ou le moteur de recherche *Google Scholar* (<https://scholar.google.com/>) qui présente des résultats intéressants en termes de couverture.^{9, 10} Ces outils sont toutefois développés en anglais : ils tiennent peu compte des spécificités des autres langues et ne disposent pas toujours d'interfaces adaptées aux utilisateurs non anglophones. En Amérique latine, la base de données bibliographiques Lilacs (*Literatura Latino-Americana e do Caribe em Ciências da Saúde* [<http://lilacs.bvsalud.org/en/>])¹¹ existe en trois langues : espagnol, portugais et anglais. En France, la principale base de données bibliographiques en sciences, technologies et médecine PAS-CAL, développée par l'Institut national de l'information scientifique et technique (INIST) depuis 1973 n'est plus mise à jour depuis fin 2014.¹² Il n'existe donc >>>

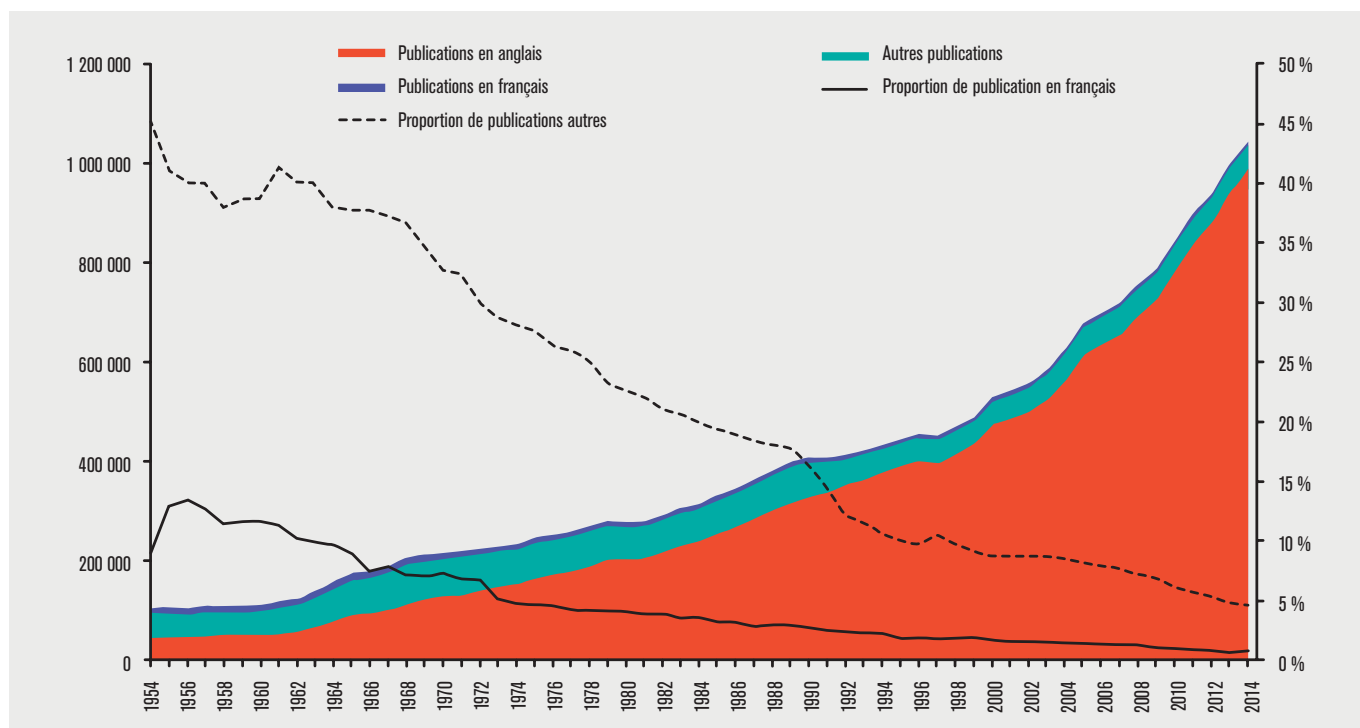


Figure 1. Nombre (échelle de gauche) et proportion (échelle de droite) de publications françaises dans PubMed, par année. Source : PubMed.

plus, à notre connaissance, de base de données bibliographiques à jour couvrant tous les domaines de la santé et prenant en compte les spécificités du français.

Pour donner plus de visibilité, de vitalité et rendre accessible la littérature médicale et paramédicale en français, le projet *Base de données bibliographiques en français* (BDBfr), financé par l'Agence nationale de la recherche (ANR) dans le programme *Technologies pour la santé* (TecSan), vise à créer une base de données bibliographiques en santé et en français, disponible gratuitement à travers un site Internet « LiSSa, Littérature scientifique en santé » (<http://www.lissa.fr>). Ce projet rassemble deux équipes académiques spécialisées en e-santé (le service d'informatique biomédicale du CHU de Rouen et le centre d'investigation clinique-innovation technologique [CIC-IT] de Lille) et trois acteurs industriels (un leader de l'édition scientifique, Elsevier Masson, et deux sociétés privées françaises innovantes, Alicante et SenseGates).

D'où viennent les données bibliographiques ?

Les données bibliographiques de LiSSa proviennent soit directement des éditeurs, soit de la base de données PubMed produite par la *National Library of Medicine* (NLM) américaine. La NLM met gratuitement à disposition sur le web ses données bibliographiques déposées dans le domaine public.¹³ On y trouve des données en langue française publiées depuis 1912 jusqu'à

Les auteurs déclarent n'avoir aucun lien d'intérêts.

Le projet LiSSa a été partiellement financé par le programme TecSan de l'ANR, sous le nom de projet BDBfr, Base de données bibliographiques en français (n° ANR-14-CE17-0020). LiSSa est disponible gratuitement : <http://www.lissa.fr>.

nos jours. Le nombre de ces publications par année décline chaque année depuis 1970 ($n = 15\,924$), jusqu'à nos jours ($n = 8\,104$). Pour chaque article, peu d'informations en langue française sont disponibles puisque le résumé et les mots-clés (*medical subject heading* [MeSH]) décrivant l'article sont exclusivement en anglais. Un programme spécifique a été développé pour permettre une mise à jour quotidienne de LiSSa avec les données issues de PubMed.

L'éditeur de journaux médicaux scientifiques proposant le plus de titres en langue française est aujourd'hui Elsevier Masson,¹⁴ partenaire du projet. Les données bibliographiques d'Elsevier Masson disponibles au format numérique sont majoritairement postérieures à 2000. Certains des articles sont indexés dans PubMed et les données d'Elsevier Masson permettent un enrichissement de leur description (affiliations détaillées de tous les auteurs, résumés en français). Certains ne sont pas indexés dans PubMed et constituent de nouvelles entrées. Le partenariat avec Elsevier Masson nous a notamment permis d'indexer dans LiSSa plusieurs revues à destination des professions paramédicales, comme *Kinésithérapie*, *la revue* ou *L'Aide-soignante*.

Le Collège national des généralistes enseignants propose les données numériques de sa revue *exercer*. Cet organisme a été contacté afin de rendre ces données accessibles depuis LiSSa car les médecins généralistes sont considérés par le consortium comme un public

pouvant être particulièrement intéressé par cet outil.

Un important travail manuel sur ces bases de données est nécessaire pour harmoniser les métadonnées afin de les intégrer aisément dans notre modèle de données.

MODÈLE DE DONNÉES ET MOTEUR DE RECHERCHE

Le service d'informatique biomédicale du CHU de Rouen développe depuis de nombreuses années un catalogue de ressources¹ et un moteur de recherche associé.¹⁵ Ces derniers ont constitué les éléments de base pour la création de LiSSa.

Le modèle de données du catalogue CISMef était initialement prévu pour décrire des ressources Internet.¹⁶ Il a dû être modifié et enrichi, de manière à pouvoir gérer des articles scientifiques et de formation continue contenus dans nos deux principales sources de données initiales : l'ensemble des références bibliographiques en langue française de PubMed (librement disponible) et d'Elsevier Masson. Nous nous sommes appuyés sur le *Dublin Core*¹⁷ (<http://dublincore.org/>) et le *Learning Object Metadata* (LOM) [IEEE 1484.12.1 – 2002] à cette fin.

Le modèle de données sur lequel reposait le moteur de recherche sémantique du catalogue CISMef ayant été modifié, nous avons dû adapter ce dernier. Il a fallu le transformer en moteur générique capable non plus d'interroger une base de données intégrant uniquement des ressources du Web, mais n'importe quelle base de données contenant tout type de documents, en l'occurrence, pour LiSSa, des citations d'articles.¹⁸

ADAPTATIONS

Les premiers développements de LiSSa ont reposé sur l'interface de CISMef. Quelques fonctionnalités ont toutefois été ajoutées.

L'analyse des métadonnées fournies par PubMed et Elsevier Masson a d'abord révélé que les métadonnées de langage de PubMed étaient parfois inexactes, avec des articles qui ne sont pas en français et pourtant étiquetés comme tels. Une première analyse a permis d'éliminer quelques dizaines d'erreurs (par exemple, certaines revues publient des articles dans diverses langues et l'étiquetage des langues est alors imprécis), mais il n'est pas possible d'être exhaustif.

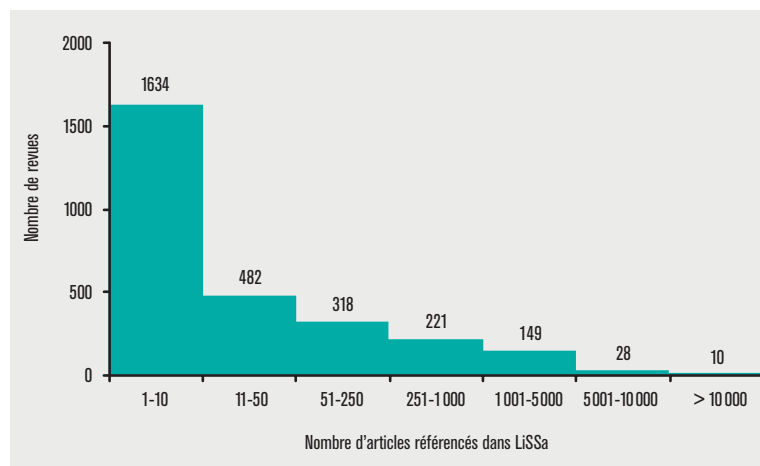


Figure 2. Distribution des revues selon le nombre de références en français intégrées dans LiSSa.

Cette analyse a également permis d'adapter le modèle et le moteur de recherche avec la création de 12 nouveaux types d'informations (ou métadonnées) compatibles avec les standards existants (LOM et *Dublin Core*). Plus de 600 000 références en français publiées dans 2769 journaux, dont *La Revue du Praticien*, ont été extraites de PubMed et intégrées dans LiSSa. PubMed ne dispose toutefois d'aucun résumé en français. Seule une minorité de ces journaux publie régulièrement en français tandis qu'une majorité publie des articles francophones de façon ponctuelle (fig. 2). Elsevier Masson a fourni des données sur plus de 200 000 références publiées dans 125 revues médicales en langue française. Au total, LiSSa compte 832 446 références d'articles. L'intersection entre le corpus PubMed francophone et les ressources Elsevier-Masson est de 51 413 références. Elsevier Masson a fourni 80 535 résumés, dont 32 310 concernent des articles référencés dans PubMed. La généricité du modèle de données nous a permis d'intégrer sans difficulté les références publiées dans *exerçer*, soit 704 références. Une description de l'ensemble des références est disponible dans le **tableau ci-contre**, ainsi qu'un détail sur les données postérieures à 2000.

Il est possible d'exporter des références vers des logiciels de gestion bibliographique et d'utiliser des filtres dynamiques (facettes) pour préciser les résultats. >>>

TABEAU. DESCRIPTION DU CONTENU DE LISSA

	Total (y compris <i>exerçer</i>)	PubMed total	Elsevier Masson (à l'exclusion des données déjà présentes dans PubMed)
Références	832 446	677 259	154 483
Références postérieures à 2000	265 195	129 014	135 477
Revue	2 842	2 769	125
Revue en cours de publication	166	100	65

LITTÉRATURE SCIENTIFIQUE

EN DÉBAT

The screenshot displays the LiSSa search results page for the query 'asthme'. The interface includes a search bar at the top with the text 'asthme' and a green checkmark. Below the search bar, it indicates 'concept(s) identifié(s) : 0804493 - asthme asthme'. The main results area shows '4783 entrées trouvées en 0,88 s' and a list of search results. On the left, there are filters for 'Type de publication' (with options like 'article de périodique', 'revue de la littérature', 'présentations de cas', 'étude comparative') and 'Année de publication' (with years 2015, 2014, 2013, 2012). The search results list includes titles, authors, and journals. For example, the first result is 'Immunothérapie et asthme allergique : analyse critique des mécanismes immunologiques à l'origine de l'induction de tolérance' by Lungoci E, Terrade J-E, Jalles C, Devouassoux G, Nicolas J-F, Berard F, published in 'Revue Française d'Allergologie' in 2015. The second result is 'Sous-diagnostic de l'asthme chez les enfants en classe de CE2' by Walus I, Richard G, Laquerrière B, Perucca M, Tuveri R, Einbinder V, Muller B, Beydon N, published in 'Archives de pédiatrie' in 2015. The third result is 'L'organisation spatiale de touffes de cils gouverne le transport de mucus dans les voies aériennes. Application à l'asthme sévère' by Khelloufi M-K, Garulli C, Viallat A, Chanez P, Gras D, published in 'Revue des maladies respiratoires' in 2015. The fourth result is 'Rôle du stress mécanique sur l'expression des protéines de jonction épithéliale dans l'asthme' by Cornetto MA, Arumugam G, Fernandes J, Pretolani M, Létuvé S, Taillé C, published in 'Revue des maladies respiratoires' in 2015. The interface also includes a 'Se connecter' link, a 'Envoyer vers' dropdown, and a 'Réponses' dropdown set to '20'. At the bottom, there is a pagination bar showing '1 2 3 4 5 6 7 8 9 10 ... 240' and a footer with 'v3.0 Contact - © 2015 CHU de Rouen - CISMef - Powered by Vaadin'.

Plus de 800 000 références en français

Ce travail a permis de mettre à la disposition du grand public une base de données bibliographiques intégrale en français, LiSSa (fig. 3). Cette base de données rassemble plus de 800 000 références bibliographiques, ce qui en fait, à notre connaissance, la base de données bibliographiques en langue française du domaine de la santé la plus riche.

FORCES ET FAIBLESSES

Le contenu de LiSSa est sa principale force : plus de références que les standards du domaine (à l'exclusion de *Google Scholar*, dont la ligne éditoriale est difficile à appréhender), des références plus riches avec plus de métadonnées en français (résumé, affiliation des auteurs...), une couverture plus large que la biologie/santé, avec l'inclusion de ressources à destination des professions paramédicales. L'ensemble étant rédigé en français est plus accessible pour les professionnels francophones et plus à même de répondre à leurs besoins d'information. Quoique cette base de données s'adresse en particulier aux médecins, généralistes⁴ ou pas, qui lisent de la documentation en français plutôt qu'en anglais, elle s'adresse également aux autres professionnels de santé puisque LiSSa intègre des revues spécialisées paramédicales qui, pour la plupart, ne sont pas sélectionnées par Medline.

Cette force de LiSSa, du contenu en français pour des professionnels francophones, constitue également sa principale faiblesse. Les dernières avancées de la science sont publiées en anglais avant d'être un jour, éventuelle-

ment, publiées en français. Il est néanmoins très facile de transposer la requête effectuée dans LiSSa à PubMed, ouvrant ainsi l'accès à ce type d'articles. Par ailleurs, un des objectifs du projet est de participer au dynamisme des publications en langue française en leur offrant une visibilité qui aujourd'hui leur fait défaut. Ce dynamisme ne permettra jamais au français de (re)devenir une langue majeure de communication scientifique, mais à un niveau plus pragmatique, pour les professionnels exerçant en français, on peut espérer que les revues de synthèse se fassent plus régulièrement en langue française et offrent une meilleure couverture de l'exercice médical et paramédical.

PERSPECTIVES

Une évaluation ergonomique a été réalisée et a permis d'aboutir à la formulation de recommandations. Leur implémentation permettra, à terme, de garantir l'utilisabilité de LiSSa, afin de faciliter sa prise en main et son appropriation. La mise en place d'un info-bouton¹⁹ permettant de transposer les requêtes à d'autres sources d'information, comme CISMef (ressources web gratuites) ou PubMed (littérature anglophone), va voir le jour très prochainement.

Par ailleurs, LiSSa doit être étendue dans trois dimensions dans les deux prochaines années :

- intégrer des critères bibliométriques en libre accès (comme Scimago²⁰, par exemple). En effet, peu de revues en langue française sont prises en compte lors des évaluations bibliométriques des professionnels de santé.

LiSSa pourrait participer, à terme, à l'élargissement du référencement des publications ;

– permettre une recherche d'informations, en premier lieu sur les résumés en français mais également sur le texte intégral des articles en libre accès ;

– intégrer les revues en langue française des autres éditeurs scientifiques, en gardant à l'esprit qu'Elsevier Masson couvre environ deux tiers des revues scientifiques de santé en français. À ce titre, l'éditeur Global Média Santé devrait bientôt nous fournir des données concernant ses revues : cela permettra, d'une part, d'enrichir les références de *La Revue du Praticien*, déjà intégrées par PubMed, avec notamment les résumés en français et, d'autre part, de référencer *La Revue du Praticien Médecine générale*, *Le Concours médical*, *La Revue française du dommage corporel* et l'édition française du *Journal of Clinical Oncology*.

Une collaboration avec Lilacs,¹¹ l'équivalent sud-américain de LiSSa, a déjà été engagée. L'extension du MeSH développée par BIREME, le *Descriptores en Ciencias de la Salud* (DeCS),²¹ sera utilisée pour l'indexation des articles dans LiSSa... et dans CISMef¹ par la même occasion. Cette terminologie a donc été intégrée dans notre serveur terminologique inter-lingue HeTOP (<http://www.hetop.eu/>) et sera prochainement traduite en français.

En septembre 2015, LiSSa a été intégrée à la formation des étudiants en médecine de la faculté de Rouen au niveau de la deuxième année d'études et du troisième cycle de médecine générale. Une action spécifique de formation a été menée dans les départements de médecine générale de Lille et de Nice, ainsi qu'à la Cité de la santé et dans une école de kinésithérapie de Berck pour les autres professionnels de la santé, qui seront sans doute plus intéressés par les 72 nouvelles revues intégrées dans LiSSa

jusque-là non sélectionnées par Medline. L'effort de formation dans les départements français de médecine générale sera poursuivi dans les années à venir. Nous espérons mettre en place une formation dans la moitié des facultés de médecine françaises avant la fin du projet (fin 2017). Une collaboration avec l'Agence universitaire de la francophonie pourrait permettre une extension à l'ensemble des pays francophones. 

Remerciements

Nous remercions vivement Laetitia Rollin, Ferdinand Dhombres, Séverine Loiseau, Laura Douze, Romaric Marcilly, Olivier Chabot, Philip Woodall, David Delerue, Dominique Dutoit, Badiisse Dahamna, Jean-François Gehanno, membres du consortium BDBfr, pour leur participation à la création de LiSSa.

RÉFÉRENCES

- Darmoni SJ, Leroy JP, Baudic F, Douyère M, Piot J, Thirion B. CISMef: a structured health resource guide. *Methods Inf Med* 2000;39:30-5.
- Vermandel M. IRBM, une nouvelle ère ? *IRBM* 2012;33:241-2.
- Plaud B, Ashenoune K, Beaulieu P, Beloeil H, Charbit B, Constantin JM, et al. Ce n'est qu'un au revoir. *Ann Fr Anesth Réanim* 2014;33:611-4.
- Bras PL, Ricordeau P, Roussille B, Saintoyant V. L'information des médecins généralistes sur le médicament. www.ladocumentationfrancaise.fr ou <http://bit.ly/2c7Zs7q>
- Chneineiss H. À quoi sert donc une revue scientifique en langue française ? *Med Sci* 2014;30:7-8.
- Dee CR. The development of the Medical Literature Analysis and Retrieval System (MEDLARS). *J Med Libr Assoc* 2007;95:416-25.
- Jamali J, Salehi-Marzjarani M, Ayatollahi SM. Factors affecting journal quality indicator in Scopus (SCImago Journal Rank) in obstetrics and gynecology journals: a longitudinal study (1999-2013). *Acta Inform Med* 2014;22:385-8.
- Levay P, Ainsworth N, Kettle R, Morgan A. Identifying evidence for public health guidance: a comparison of citation searching with Web of Science and Google Scholar. *Res Synth Methods* 2016;7:34-45.
- Henderson J. Google Scholar: a source for clinicians? *CMAJ* 2005;172:1549-50.
- Gehanno JF, Rollin L, Darmoni S. Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Med Inform Decis Mak* 2013;13:7.
- Moreira MA, Lustosa AM, Dutra F, Barros EO, Batista JB, Duarte MC. Public humanization policies: integrative literature review. *Cien Saude Colet* 2015;20:3231-42.
- <http://www.inist.fr/?Francis-et-Pascal-en-libre-acces>
- <http://www.ncbi.nlm.nih.gov/home/about/policies.shtml>
- <http://www.speps.pro/index.php/les-adherents/les-editeurs>
- Darmoni SJ, Thirion B, Leroy JP, et al. DocCISMEF: a search tool based on «encapsulated» MeSH thesaurus. *Stud Health Technol Inform* 2001;84:314-8.
- Thirion B, Loosli G, Douyère M, Darmoni SJ. Metadata element set in a quality-controlled subject gateway: a step to a health semantic Web. *Stud Health Technol Inform* 2003;95:707-12.
- Weibel SL, Koch T. The Dublin Core Metadata Initiative. *D-Lib Magazine* 2000. <http://www.dlib.org/dlib/december00/weibel/12weibel.html>
- Griffon N, Schuers M, Soualmia LF, et al. A search engine to access PubMed monolingual subsets: proof of concept and evaluation in French. *J Med Internet Res* 2014;16:e271.
- Darmoni SJ, Pereira S, Névéal A, et al. French Infobutton: an academic and... business perspective. *AMIA Annu Symp Proc* 2008:920.
- Falagas ME, Kouranos VD, Arencibia-Jorge R, Karageorgopoulos DE. Comparison of SCImago journal rank indicator with journal impact factor. *FASEB J* 2008;22:2623-8.
- Pereira TA, Montero EF. DeCS terminology and the new rules on orthography of Portuguese language: guidelines for an update. *Acta Cir Bras* 2012;27:509-14.



Littérature Scientifique en Santé (LiSSa) : une alternative à l'anglais ?

Nicolas Griffon^{1,2}, Matthieu Schuers^{1,2}, Stefan J. Darmoni^{1,2}

1. CHU de Rouen, LITIS-TIBS EA 4108, service d'informatique biomédicale, 76031 Rouen cedex, France
2. LIMICS, Inserm, U1142, 75006 Paris, France

Correspondance :

Nicolas Griffon, CHU de Rouen, LITIS-TIBS EA 4108, service d'informatique biomédicale, 76031 Rouen cedex, France.
nicolas.griffon@chu-rouen.fr

LiSSa: An alternative in French to browse health scientific literature ?

Internet permet aujourd'hui aux professionnels de santé francophones d'accéder à une grande quantité d'informations en français. De nombreux services de documentation en langue française [1-4] recensent et permettent d'accéder à la « littérature grise » francophone (à savoir la production de toutes les instances du gouvernement, de l'enseignement et la recherche publique, du commerce et de l'industrie, qui n'est pas contrôlée par l'édition commerciale), et notamment aux recommandations de bonnes pratiques ou aux thèses. À l'inverse, pour les articles scientifiques, la plupart des services sont en anglais... même lorsque l'on recherche des références francophones. Cela peut sembler logique au vu de la prédominance de l'anglais dans les sciences de la santé, mais est ce judicieux ? Pubmed, Google Scholar, PASCAL, Web Of Science... Chaque base de données a des limites linguistiques qui rendent son utilisation délicate pour les professionnels de santé francophones. PubMed est intégralement en anglais : la plupart des informations sur les références (titre, résumé, mots clés...), aussi appelées métadonnées, ne sont disponibles que dans cette langue et le moteur de recherche n'est pas adapté au français. La ligne éditoriale de Google Scholar n'est pas explicite et les filtres par langue sont parfois d'une efficacité discutable. La base de données PASCAL n'est plus mise à jour depuis fin 2014 [5]. Web Of Science intègre de nombreuses sources de données mais souffre des mêmes défauts que PubMed. En définitive, il n'existe pas, à notre connaissance, de base de données bibliographiques à jour, couvrant tous les domaines de la santé et prenant en compte les spécificités de la langue française. La visibilité et l'impact des publications en français, incluses dans les bases existantes sont ainsi forcément limités. Leur valorisation, qui repose sur le facteur d'impact des revues, tant en termes de financement de la recherche par les institutions, qu'en termes de carrière pour les auteurs, s'en

trouve amoindrie, accélérant encore la disparition des publications en français des bases de données.

Pourtant, lire en français est plus naturel que lire en anglais pour plusieurs dizaines de millions d'individus [6] ; et, malgré sa prédominance, l'anglais est un frein, voire un obstacle à la lecture pour de nombreux professionnels de santé [7] français ou francophones. Les articles de synthèse en français restent plébiscités par le lectorat francophone [8] et permettent une acquisition des connaissances plus aisée. Les publications en français ont encore toute leur utilité ; leur disparition sera sans doute préjudiciable à la mise à jour des connaissances et donc, *in fine*, à la qualité des soins.

Alors que l'accumulation continue des connaissances rend toute recherche d'information de plus en plus complexe, il est paradoxal d'y ajouter une difficulté : celle de la langue.

C'est de ces constats et de cette inquiétude, partagée par l'Académie nationale de médecine [9] qu'est née l'idée de créer une base de données bibliographiques améliorant la visibilité, l'accessibilité et le dynamisme de la littérature médicale et paramédicale en français. L'acte de naissance du site Internet « LiSSa, Littérature Scientifique en Santé » (URL : <http://www.lissa.fr/>) a été signé par l'Agence nationale de la recherche en 2015, quand elle accepte de financer, par le programme Technologies pour la Santé (TecSan), le projet de base de données bibliographiques en français (n° ANR-14-CE17-0020). Ce projet réunit le service d'informatique biomédicale du CHU de Rouen, le CIC-IT de Lille et trois industriels : Elsevier Masson (EM), Alicante et Sensgates.

Bien qu'en cours de développement, ce site rassemble déjà les références bibliographiques des publications en français présentes dans PubMed, une grande partie des publications santé d'EM en français et des articles publiés dans la revue *Exercer*. À ce jour, LiSSa englobe 832 446 références, parmi lesquelles plus de 260 000 postérieures à l'an 2000 et 81 239 disposant d'un résumé en français (inexistant dans PubMed). Plus de 600 000 références en français, publiées dans 2769 journaux, ont été extraites de PubMed et intégrées dans LiSSa. Elsevier Masson a fourni des données sur plus de 200 000 références publiées dans 125 revues médicales en langue française, parmi lesquelles près de 70 non indexées dans PubMed.

L'intersection entre le corpus PubMed francophone et les ressources EM est de 154 483 références. Un effort particulier a été porté à la fusion des informations concernant les références de cette intersection, avec un effet positif sur la richesse des métadonnées et sur le nombre de « doublons ».

Ce travail a permis de mettre à disposition, de façon libre et gratuite, la base de données bibliographiques rédigées en langue française sans doute la plus riche existant actuellement dans le domaine de la santé. Les retours obtenus lors de l'évaluation ergonomique semblent indiquer que LiSSa est un outil prometteur pour les médecins.

Le contenu de LiSSa est sa principale force : plus de références que les standards du domaine (à l'exclusion de Google Scholar, déjà évoqué ci-dessus) ; des références plus riches avec plus de métadonnées en français (résumé, lien vers le texte intégral, gratuit ou pas...) ; et une couverture plus large que la biologie/santé, avec l'inclusion de ressources à destination des professions paramédicales, rarement considérées par PubMed. Ainsi, quoique cette base de données s'adresse surtout aux médecins qui lisent de la documentation en français plutôt qu'en anglais, elle concerne aussi les autres professionnels de santé. L'ensemble rédigé en langue française est plus accessible aux francophones et plus à même de répondre à leurs besoins d'information. Nous avons pris contact avec les principaux éditeurs de journaux en français, outre EM, pour étendre encore notre couverture et pallier nos manques (*Annales françaises de médecine d'urgence* par exemple).

Des actions spécifiques de formation ont déjà été menées à la Cité de la Santé, dans une école de kinésithérapie de Berck et dans 3 départements de médecine générale (Rouen, Lille et Nice). Le projet a pour objectif d'étendre ces formations à la moitié des facultés de médecine françaises avant fin 2017. Nous visons enfin une extension à l'ensemble des pays francophones via une collaboration avec l'Agence universitaire de la francophonie.

À suivre !

Déclaration d'intérêts : les auteurs n'ont pas précisé leurs éventuels liens d'intérêts.

Références

- [1] CISMef. [<http://www.cismef.org/>].
- [2] Recomedical. [<http://recomedical.fr/>].
- [3] Bibliothèque interuniversitaire de santé. [<http://www.biusante.parisdescartes.fr/>].
- [4] Bibliothèque Médicale A.F. Lemanissier. [<http://www.bmlweb.org/>].
- [5] Francis et Pascal en libre accès. INIST. [Document Web, archivé par WebCite® : <http://www.webcitation.org/6fqoCMY9y>].
- [6] Wikipédia. Article sur la francophonie. [Document Web, archivé par WebCite® : <http://www.webcitation.org/6g2z0wPMM>].
- [7] Bras PL, Ricordeau P, Roussille B, Saintoyant V. L'information des médecins généralistes sur le médicament. [Document Web, archivé par WebCite® : <http://www.webcitation.org/6fqoPoXPX>].
- [8] Chneiweiss H. À quoi sert donc une revue scientifique en langue française ? *Med Sci* 2014;30:7-8.
- [9] Fiessinger JN. Valorisation des revues médicales en langue française. Communiqué de l'Académie nationale de médecine; 2016 [Document Web, archivé par WebCite® : <http://www.webcitation.org/6kf72Bc7a>].

4. Le CRBM : Constructeur de requêtes bibliographiques médicales

4.1. Requêter dans PubMed dans sa langue maternelle : les enjeux de la création d'un outil francophone

Plus de 82% des citations présentes dans PubMed sont en anglais. Cela constitue probablement un obstacle pour de nombreuses personnes souhaitant accéder à l'information contenue dans la base de données (92). Plusieurs outils permettent actuellement aux utilisateurs non-anglophones de requêter dans PubMed dans leur langue maternelle. Citons notamment BabelMeSH (<https://babelmesh.nlm.nih.gov/>) et PICO Linguist (<https://babelmesh.nlm.nih.gov/pico.php>), développés aux Etats-Unis par l'Institut National de la Santé (*National Institute of Health*, NIH) (93). Ces deux outils permettent de requêter dans Pubmed en 12 langues pour le premier et 14 pour le second. Ils permettent également de sélectionner la ou les langues des références retrouvées. BabelMesh permet de créer des requêtes contenant un à deux termes MeSH, il ne permet pas l'utilisation des opérateurs booléens classiques. PICOLinguist permet de créer des requêtes plus complexes en utilisant la structure PICO (Patient, Intervention, Comparison, Outcome), l'utilisation d'opérateurs booléens n'est pas non plus possible. Pour les deux outils, les résultats sont présentés sur une page spécifique intégrée à l'outil, il n'existe pas de renvoi vers PubMed permettant d'accéder aux fonctions de tri ou de sélection des références disponibles au sein de la base de données.

Les résultats de notre étude qualitative ont confirmé le manque de compétences des médecins généralistes à la fonction de recherche dans les bases de données. Ils n'en maîtrisent pas les procédures spécifiques de requêtage. Il est possible que le fait que l'anglais soit souvent utilisé dans ces moteurs de recherche explique ces difficultés. Les outils existants ne nous paraissant pas répondre totalement aux besoins des utilisateurs, notamment en termes de fonctionnalités, il nous a paru intéressant de développer un

nouvel outil, facilitant le requêtage dans PubMed dans sa langue maternelle et disposant de fonctionnalités avancées.

4.2. Le constructeur de requêtes bibliographiques médicales : description de l'outil et de ses fonctionnalités

Lors de sa création, l'objectif de l'équipe du D2IM était de construire un outil facile à utiliser pour les personnes non anglophones, permettant de retrouver les mots clés MeSH pertinents pour une requête dédiée.

L'outil est basé sur les traductions du MeSH disponibles. Il permet aux utilisateurs de disposer à la fois d'une interface et d'un vocabulaire contrôlé dans leur propre langue maternelle. L'interface est actuellement disponible en français, anglais et espagnol. Le thésaurus MeSH est disponible dans huit langues : espagnol, anglais, finnois, français, italien, néerlandais et portugais.

L'utilisateur du CRBM remplit le cartouche de recherche avec du langage naturel. Grâce à la fonction d'auto-complétion, le ou les termes MeSH correspondant à sa requête lui sont proposés. A chaque terme sont associées plusieurs informations, dont la définition du terme, ses relations au sein de la terminologie et une liste hiérarchisée de ses qualificatifs. Une fois le terme adéquat choisi par l'utilisateur, celui-ci peut être ajouté à la requête. D'autres termes MeSH peuvent être sélectionnés et insérés dans la requête, à l'aide des opérateurs booléens habituels. Dès lors qu'une requête satisfaisante est obtenue par l'utilisateur, celui-ci peut la soumettre à trois bases de données : CISMeF, LiSSa et PubMed. L'outil permet de revenir en arrière en cas d'erreur, de visualiser le nombre de références obtenues par chaque requête dans chacune des trois bases de données. En plus des termes MeSH sélectionnés par l'utilisateur, chaque requête est complétée par l'ajout de synonymes, destinés à augmenter la précision et le rappel. Comme montré par Thirion et al (94), l'ajout de termes permet d'augmenter de manière significative la précision des résultats, ceci comparé à la requête par défaut construite par PubMed (54,5% versus 27%).

Le constructeur de requêtes est une application web, développée en Java grâce au framework Vaadin, connectée à quatre services. Chacun de ces services est dédié à une tâche spécifique :

- La fonction d'auto-complétion permet l'accès aux termes MeSH en rapport avec la requête,
- Les informations associées au terme MeSH sélectionné sont extraites à partir du serveur terminologique,
- Le service InfoRoute est l'application principale du constructeur de requête, elle permet de construire l'URL PubMed en utilisant les différents éléments syntaxiques de la base de données (termes MeSH, opérateurs booléens...),
- Une fonction permettant la visualisation du nombre de références obtenues avec la requête dans chacune des bases de données.

4.3. Données d'utilisation du constructeur de requêtes bibliographiques médicales

En 2016, le CRBM a été visité 13 491 fois par plus de 8 300 visiteurs différents. Près d'un million de pages y ont été lues. Environ 85% des utilisateurs sont originaires de France.

4.4. Publication

Est présenté ci-après un article relatif au constructeur de requêtes bibliographiques médicales, publié dans la revue BMC Medical Informatics and Decision Making. L'objectif de cette étude était d'évaluer, chez des internes de médecine générale français, l'impact sur la qualité des requêtes de l'utilisation de leur langue maternelle pour requêter dans PubMed.

4.5. Un constructeur de requêtes multilingue améliore la qualité des requêtes effectuées dans PubMed : un essai contrôlé randomisé.

4.5.1. Contexte

Les utilisateurs francophones de PubMed doivent faire face à plusieurs obstacles. L'un d'entre eux est probablement le maniement des mots-clés en anglais. Afin de faciliter le requêtage dans PubMed, nous avons développé un outil permettant d'interroger la base de données en français : il s'agit du Constructeur de Requêtes Bibliographiques Médicales (CRBM). Il s'agit du premier outil disponible permettant d'élaborer des requêtes complexes en français dans la base de données PubMed.

4.5.2. Objectif

Evaluer l'impact du requêtage dans PubMed dans sa langue maternelle sur la qualité des requêtes effectuées.

4.5.3. Méthode

Un essai contrôlé randomisé a été mené dans le département de médecine générale de l'université de Rouen en janvier 2015. Des internes de médecine générale dont le français était la langue maternelle ont été recrutés par mail et répartis aléatoirement en deux groupes, l'un disposant de la version française du constructeur de requêtes, l'autre de la version anglaise. Ils étaient chargés de traduire 12 questions médicales courtes (les mêmes pour tous les participants) en requêtes MeSH. Les questions étaient écrites en français pour les deux groupes. Ils reçurent en amont de l'évaluation une session de formation d'une quinzaine de minutes, en français, portant sur la recherche bibliographique en santé. Les 12 questions médicales ont été proposées par un documentaliste médical et validées par deux médecins. Elles étaient réparties selon trois niveaux de complexité, selon le nombre de termes MeSH nécessaires à la création d'une requête correcte.

Le constructeur de requêtes a été intégré dans une application web. Elle permettait de verrouiller la langue de l'interface (anglais ou français), de proposer une page différente

pour chacune des questions médicales, d'enregistrer la requête soumise et de mesurer le temps de réponse du participant à chaque question. Les participants pouvaient passer librement d'une question à l'autre et modifier leurs réponses.

Le critère de jugement principal était la qualité de la requête. Toutes les requêtes construites par les participants ont été fournies en anglais aux documentalistes pour évaluation. L'évaluation de la qualité de la requête a été effectuée à partir d'une classification existante dans la littérature. Deux documentalistes ont évalué indépendamment chacune des requêtes. Les divergences ont été résolues par consensus. Le temps passé à la construction des requêtes constituait le critère de jugement secondaire.

4.5.4. Résultats

Quarante-deux internes ont participé à cette étude. Vingt ont utilisé la version française du constructeur de requêtes, 22 la version anglaise. Au total, 492 requêtes ont pu être analysées.

Cent trente-cinq requêtes (27,4%) ont été considérées comme parfaites. Elles étaient significativement plus nombreuses dans le groupe français que dans le groupe anglais (respectivement 37,9% versus 17,9% ; $p<0,01$). Les erreurs les plus fréquentes étaient l'utilisation d'un terme MeSH manquant de précision (44,7%) et l'utilisation d'un terme MeSH non pertinent (27,4%). On retrouvait significativement plus de requêtes parfaites dans le groupe français que dans le groupe anglais pour les questions de difficulté faible (48,1% versus 31,8% ; $p=0,04$), les questions de difficulté moyenne (45,6% versus 18,4% ; $p<0,01$) et les questions de difficulté haute (20,3% versus 3,5% ; $p<0,01$).

Les participants du groupe anglais mettaient significativement plus de temps à construire chaque requête que les utilisateurs du groupe français, respectivement 194 secondes versus 128 secondes ; $p<0,01$.

Une présentation détaillée de la méthode et des résultats de cette étude figure ci-après.

RESEARCH ARTICLE

Open Access



Lost in translation? A multilingual Query Builder improves the quality of PubMed queries: a randomised controlled trial

Matthieu Schuers^{1,2,3*} , Mher Joulakian¹, Gaetan Kerdelhué¹, Léa Segas¹, Julien Grosjean¹, Stéfan J. Darmoni^{1,4} and Nicolas Griffon^{1,4}

Abstract

Background: MEDLINE is the most widely used medical bibliographic database in the world. Most of its citations are in English and this can be an obstacle for some researchers to access the information the database contains. We created a multilingual query builder to facilitate access to the PubMed subset using a language other than English. The aim of our study was to assess the impact of this multilingual query builder on the quality of PubMed queries for non-native English speaking physicians and medical researchers.

Methods: A randomised controlled study was conducted among French speaking general practice residents. We designed a multi-lingual query builder to facilitate information retrieval, based on available MeSH translations and providing users with both an interface and a controlled vocabulary in their own language. Participating residents were randomly allocated either the French or the English version of the query builder. They were asked to translate 12 short medical questions into MeSH queries. The main outcome was the quality of the query. Two librarians blind to the arm independently evaluated each query, using a modified published classification that differentiated eight types of errors.

Results: Twenty residents used the French version of the query builder and 22 used the English version. 492 queries were analysed. There were significantly more perfect queries in the French group vs. the English group (respectively 37.9% vs. 17.9%; $p < 0.01$). It took significantly more time for the members of the English group than the members of the French group to build each query, respectively 194 sec vs. 128 sec; $p < 0.01$.

Conclusions: This multi-lingual query builder is an effective tool to improve the quality of PubMed queries in particular for researchers whose first language is not English.

Keywords: Search engine, Medical information retrieval, PubMed/MEDLINE

Background

Evidence based medicine is increasingly encouraged in medical practice and decision-making, which requires evidence based on valid research. MEDLINE, created by the US National Library of Medicine (NLM), is the most widely used medical bibliographic database in the world. It is the largest component of PubMed, which is the largest free online database of biomedical journal citations

and abstracts. PubMed currently contains 26,415,890 citations from 5,650 indexed journals from 81 countries and in 60 languages. Each PubMed record is indexed with the NLM's controlled vocabulary, the Medical Subject Headings (MeSH) [1].

More than 82% of PubMed citations are in English and this can be an obstacle for some researchers to access the information the database contains [2]. Nevertheless, some tools are, in fact, available to help non-native-English speakers to access PubMed references written in their native language: i.e. BabelMeSH [3, 4], Patient, Intervention, Comparison, Outcome (PICO) Linguist [4] and LiSSa [5]. Although some of these tools have demonstrated a high

* Correspondence: matthieu.schuers@chu-rouen.fr

¹Department of Biomedical Informatics, Rouen University Hospital, Rue de Germont, 76031 Rouen Cedex, France

²TIBS, LITIS EA 4108, Rouen University, Rouen, France

Full list of author information is available at the end of the article



level of precision and coverage [5], they can only permit limited access to available evidence.

Recent research has also confirmed the lack of skills to perform a literature search among physicians: they not only are unable to master the specific querying process of medical databases but also feel uneasy in performing research [6]. The English used in the PubMed querying process might possibly explain some of the difficulties. Therefore, we were prompted to create a multilingual query builder to facilitate access to the PubMed subset using a language other than English (e.g. French, German, Spanish, or Norwegian), with an advanced multifunctional system. This practical tool relies on the MeSH translation in multiple languages to boost information retrieval.

The aim of this study was to assess the impact of a multilingual query builder on the quality of PubMed queries for physicians and medical researchers, in particular those whose first language is not English.

Methods

The multi-lingual Query Builder

Many users, including the French, have difficulties writing queries when performing research on PubMed. Therefore, we designed a multi-lingual query builder to facilitate information retrieval. Our objective was to develop an “easy-to-use” tool to select relevant MeSH terms. This tool is based on available MeSH translations, providing users with both an interface and a controlled

vocabulary in their own language. Currently, the interface has been translated into French, English and Spanish while the MeSH thesaurus is available in more than fifteen languages (e.g. Dutch, English, Finnish, French, German, Italian, Portuguese and Spanish). Users can fill the search box with natural language. The Autocomplete search function helps to find the relevant MeSH terms. Full MeSH information is available, including definitions, relations and a hierarchical qualifiers list. The selected MeSH term is added to the query. Other MeSH terms can be added, using usual Boolean operators. When a satisfactory query is obtained, the user can submit it to PubMed. The tool allows to go back and forth: the user can observe the number of results given by the query - or the results itself - and reformulate the query if needed. An example for a given query appears in Fig. 1 (English interface). The query is then completed by using a wide range of synonyms such as natural language terms, in order to maximize recall and precision. As shown by Thirion et al., adding terms to the query significantly increases the precision when compared to the default query used by PubMed (54.5% vs 27%) [7].

The query builder is a web application (written in Java thanks to the Vaadin framework) connected to four services. Each service is dedicated to a specific task: i) the autocomplete function provides the MeSH terms related to the query; ii) the terminology server retrieval collects data of the selected MeSH term; iii) the

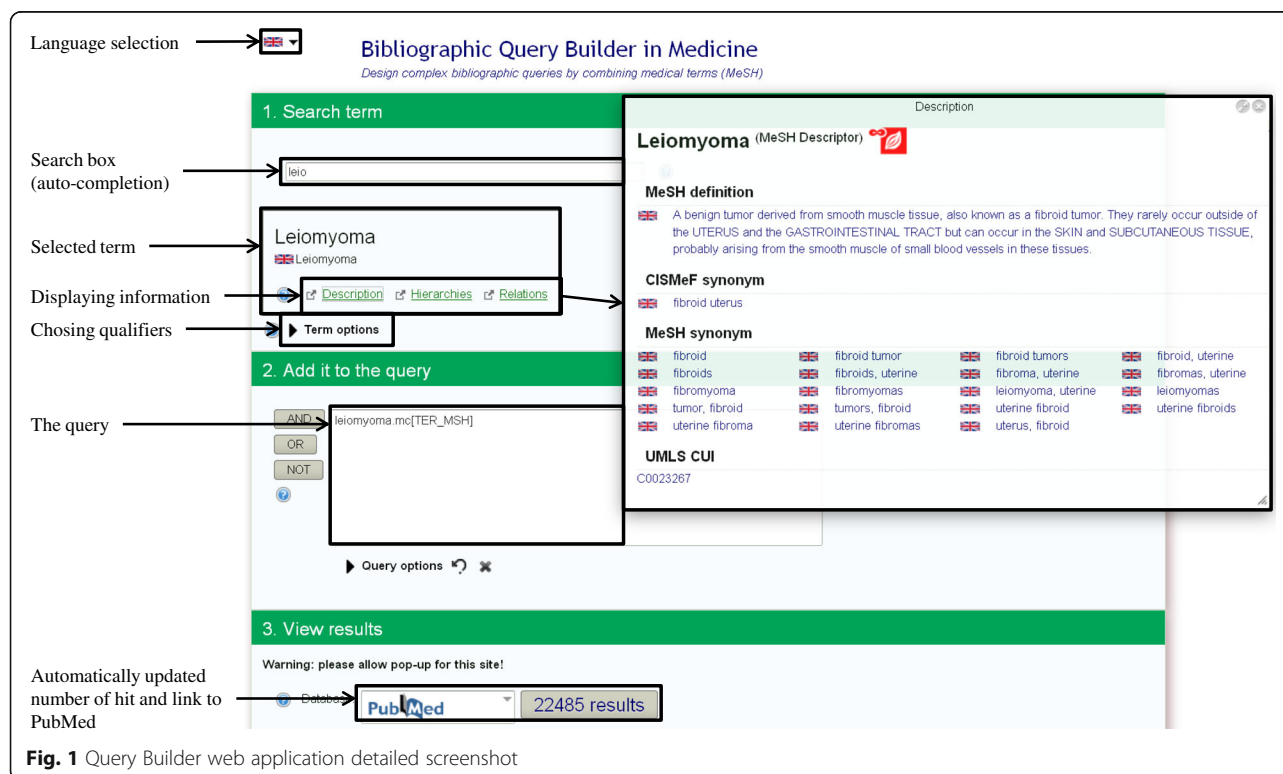


Fig. 1 Query Builder web application detailed screenshot

InfoRoute service is the main application service, it builds the PubMed URL using the advanced PubMed search syntax (search tags [8], MeSH terms, Boolean operators) and iii) a function to get the results number in PubMed for the generated query.

Recruitment and study set-up

A randomised controlled study was conducted at the Department of General Practice of Rouen University during the month of January 2015. Native French speaking general practice residents were recruited by email by one researcher (MS) and randomly allocated either the French or the English version of the query builder web tool, using a computer-generated randomisation sequence. The residents were asked to translate 12 short medical questions (the same for each participant) into MeSH queries. The questions were written in French for both groups. They received a 15 minutes training about the query builder. During the training session, the participants used the query builder in the language for which they were randomized. This training was performed in French for all participants. It focused on the different stages of bibliographic search, the description and the use of MeSH thesaurus, Boolean operators and subheadings. The evaluation took place in two adjacent rooms of the Rouen University Medical School, at the same time. Residents allocated to the first room had access to the French version of the query builder, residents allocated to the second one had access to the English version.

Twelve short medical questions were designed with three levels of complexity. The clinical questions were drafted by an experienced medical librarian (GK). They were chosen arbitrarily in the logs of a document database developed in the Rouen University hospital (<http://www.chu-rouen.fr/cismef/>). They were then validated by two physicians (MS and NG) for medical relevance. The level of complexity was decided in regard to the number of MeSH terms (descriptors or subheadings) required for a correct query. Low difficulty questions required two terms. Medium difficulty questions required three terms. High difficulty questions required four terms. The difficulty to express an idea with MeSH terms was also taken into account for one question (salty taste in the mouth). The list of short medical questions is shown in Table 1.

For evaluation purposes, the query builder was embedded in a light web application. It allowed investigators to lock the interface language (English vs. French), to propose one different page for each short medical question, record the submitted query and the overall response time. Participants were free to navigate between each medical question and change their queries.

Table 1 Short medical questions (EN/FR) and queries considered as correct

Low difficulty questions	
Fibroid uterus spontaneous rupture/ Rupture spontanée d'un fibrome utérin	Leiomyoma AND rupture, spontaneous
Alopecia areata prevention/ Prévention des pelades	Alopecia areata/prevention and control
Vitamin D determination in blood/ Dosage de la vitamine D dans le sang	Vitamin D/blood; vitamin D/analysis OR vitamin D/blood; vitamin D/blood OR (vitamin D AND blood chemical analysis)
Sarcopenia for over 65 years old patients/Sarcopénie chez les patients de plus de 65 ans	Sarcopenia AND aged
Medium difficulty questions	
Vaccination induced pain in infant/ Douleur au cours de la vaccination des nourrissons	Infant AND Pain AND vaccination; infant AND (Pain OR pain measurement OR pain management) AND vaccination
Guidelines for breast cancer treatment/Recommandations sur le traitement du cancer du sein	Breast neoplasms/therapy AND practice guidelines as topic; breast neoplasms/therapy AND practice guideline
Asthma epidemiology in USA/ Épidémiologie de l'asthme aux États-Unis	Asthma/epidemiology AND united states; (asthma/epidemiology OR (asthma AND epidemiology)) AND united states
Screening for uterine cervical neoplasm/Dépistage du cancer du col de l'utérus	Mass screening AND uterine cervical neoplasms/prevention and control; mass screening AND uterine cervical neoplasms/diagnosis;
High difficulty questions	
Salty taste in the mouth/Goût sale dans la bouche	Sodium chloride AND dysgeusia; sodium chloride AND taste disorders; sodium chloride AND taste perception
Allopurinol cutaneous side effect/ Effets secondaires cutanés de l'allopurinol	Allopurinol/adverse effect AND skin diseases/chemically induced; allopurinol/adverse effect AND (skin diseases/chemically induced OR skin manifestations/chemically induced)
Glucocorticoids effects on asthmatic patient's growth/Impact des glucocorticoïdes sur la croissance du patient asthmatique	((Glucocorticoids/adverse effects) AND (growth/drug effects OR growth disorders/chemically induced) AND asthma/drug therapy; ((glucocorticoids/adverse effects) AND (growth/drug effects OR growth disorders/etiology) AND asthma/drug therapy
Antibiotics dosage for overweight or obese patient/Posologie des antibiotiques chez le patient en surpoids ou obèse	Anti-bacterial agents/administration and dosage AND (Obesity OR overweight); (anti-bacterial agents/administration and dosage OR (drug dosage calculations AND anti-bacterial agents)) AND (Obesity OR overweight)

Questions were presented in French to all the participants. Queries only contain MeSH terms and/or MeSH terms/subheadings, except "practice guideline" which is a publication type

Outcomes and statistical analysis

The main outcome was the quality of the query. The expected queries were built by a team of medical librarians and are available in Table 1. Both groups delivered queries in English and these queries were then submitted to the librarians for evaluation. The query quality was assessed according to a modified published classification [9] that differentiated eight types of errors: irrelevant MeSH term, over-specification (descriptors or subheadings), under-specification (descriptors or subheadings), incorrect operator and syntax error. The description of these types of errors is shown in Table 2. Queries may present more than one mistake of each kind. If a query had two irrelevant MeSH terms and one under-specified qualifier, three errors were counted. Queries without any errors were considered as perfect. Two librarians blinded to the arm independently evaluated each query and resolved discrepancies by consensus. Proportions were compared using Fisher's test. Analyses were stratified according to the difficulty level.

The time spent in building queries was measured as a secondary outcome. It was measured by the web form from the reading of the clinical question to the submission of the final MeSH query by the end-user. Times were compared using the Mann-Whitney test.

For an expected improvement of 15% in the group 'native language' vs. 'English language', from 25 to 40%, 17 end-users per group were required ($\alpha = 0.05$, $\beta = 0.1$, $\text{var} = pq/n$). All statistical tests were performed with R 3.0.2 software.

Results

Forty-two residents participated in this study, each of them had at least one year of experience. Twenty

residents used the French version of the query builder and 22 used the English version. The mean age was 26.9 years old in the French group and 26.7 years old in the English group ($p = 0.58$). The proportion of male was 35% in the French group and 45.5% in the English group ($p = 0.54$). After the exclusion of 11 non-responses and one obvious mistake (query related to a previous question), 492 queries were analysed. Please see flow chart in Fig. 2.

Queries based on low difficulty questions contained 0.83 error on average [0.71–0.96], queries based on medium difficulty questions 0.92 error [0.80–1.03] and queries based on high difficulty questions 1.69 errors [1.53–1.85] ($p < 0.01$).

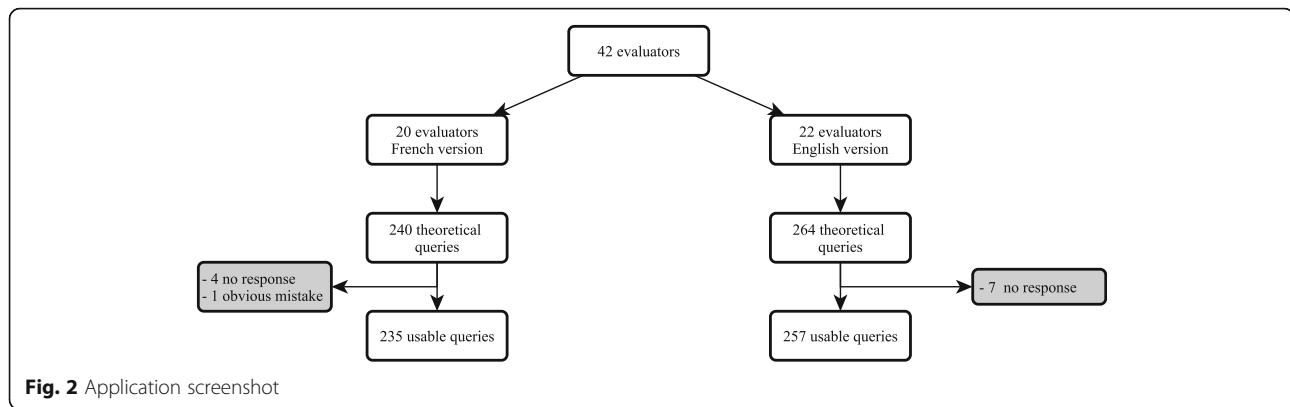
One hundred and thirty-five queries (27.4%) were considered as perfect. There were significantly more perfect queries in the French group vs. the English group (respectively 37.9% vs. 17.9%; $p < 0.01$). The most frequent errors were the use of an underspecified MeSH term (44.7%) or an irrelevant MeSH term (27.4%). Members of the French group performed significantly better for these two kinds of mistake, respectively 35.7% vs 52.9% ($p < 0.01$) and 22.1% vs. 32.3% ($p = 0.01$). No differences were found between the two groups concerning over-specification errors, the use of incorrect operator or syntax errors. See Table 2 for detailed results.

There were significantly more perfect queries in the French group vs. the English group for low difficulty questions (48.1% vs 31.8%; $p = 0.04$), medium difficulty questions (45.6% vs 18.4%; $p < 0.01$) and high difficulty questions (20.3% vs 3.5%; $p < 0.01$).

The fastest evaluator spent 85 seconds on average for each query, the slowest 260 seconds. It took significantly more time for the English participants than the French

Table 2 Summary of main results

Error type and its description (from Vanopstal et al)	French version n (%)	English version n (%)	<i>p</i>
Perfect queries	89 (37.9)	46 (17.9)	<0.01
Irrelevant MeSH term	52 (22.1)	83 (32.3)	<0.01
Query contains at least one incorrect MeSH term			
Over-specification (descriptors)	28 (11.9)	39 (15.2)	0.3
Query contains at least one MeSH term that is too narrow			
Over-specification (qualifiers)	17 (7.2)	17 (6.6)	0.8
Under-specification (descriptors)	41 (17.4)	87 (33.9)	<0.01
Query contains at least one MeSH term that is too broad			
Under-specification (qualifiers)	48 (20.4)	72 (28)	0.05
Incorrect operator	1 (0.4)	1 (0.4)	1
Misuse of "AND" or "OR"			
Syntax error	2 (0.9)	4 (1.6)	0.7
Query contains unmatched brackets or quotes, or truncated words			



participants to build each query, respectively 194 sec vs. 128 sec; $p < 0.01$ (Fig. 3).

Discussion

Summary of main results

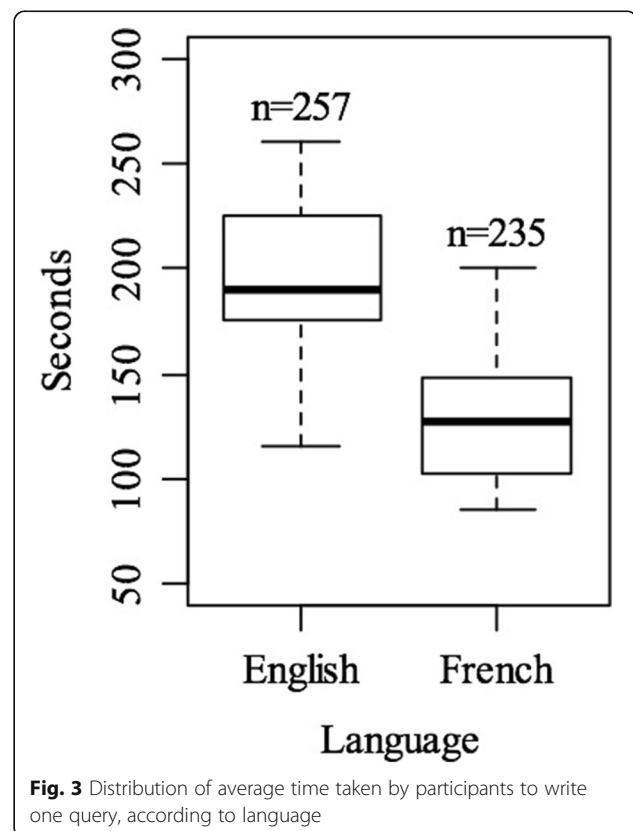
Our research findings show that a multi-lingual query builder to access PubMed could be a useful tool in research and clinical practice for non-native English speakers. Participants querying in their first language built twice more perfect queries than participants querying in English. These results were found for low, medium and high difficulty questions. The impact of querying in first language increased with the level of difficulty. Participants querying in their first language took less time to build each query than participants querying in English.

Discussion of the main results

Many barriers to query building and information retrieval among healthcare professionals have previously been identified in the literature. Currently, the most reported obstacles are: (i) the amount of time required to find information, (ii) difficulties in reformulating the original question and finding an optimal search strategy, (iii) lack of a good source of information, uncertainty as to whether all relevant information has been found and (iiii) inadequate synthesis of any pieces of evidence into a clinically useful approach [10, 11]. The literature also shows that physicians, and especially primary care doctors, express a need for database training, regardless of their first language [6, 12]. Physicians' difficulties in building search queries are well known. In 2007, a web log analysis was undertaken in a meta-search engine covering 150 health resources and a variety of guidelines. It showed that most queries were built using a single search term and no Boolean operator [13]. A similar study was conducted on PubMed queries. If PubMed queries had a median of three terms, only 11% of them contained Boolean operators [14]. Many factors can influence the

physicians' ability to build relevant search queries, including the level of English skills [9, 15]. These findings suggest that our query builder may be of significant value for non-native English speaking healthcare professionals.

As previously mentioned, our data suggest that the impact of using the first language increases with the complexity of clinical questions. Complexity appears to play a key role in physicians' difficulties in information retrieval. They fail to master the use of Boolean operators and, when dealing with complex clinical questions,



GP trainees tend to refer to their colleagues more than electronic sources [13, 16].

Participants querying in their first language took less time to build each query than participants querying in English. This appears to be an important finding, as time constraints are always cited as a major obstacle when seeking information and may improve the PubMed's use as searches with PubMed are not as frequent as searches with Google or UpToDate [10, 17].

Nevertheless, only one-third of the queries were considered perfect, even among participants querying in their first language. Irrelevant MeSH terms and the lack of specification in descriptors and subheadings lead to poor precision and recall [9]. Querying in first language will not solve all the problems faced by researchers and physicians, especially as the overwhelming majority of PubMed references remain in English. Some research tools already provide an automatic translation of a biomedical text, including titles and abstracts and using the MEDLINE database [18]. Other tools increase the information retrieval task performance by allowing non-native-English speakers to access PubMed references written in their native language: BabelMeSH [3, 4], Patient, Intervention, Comparison, Outcome (PICO) Linguist [4] and LiSSa [5]. According to Gagnon et al., educational meetings currently seem to be the only type of interventions showing a significant positive effect on clinical information retrieval technologies adoption by healthcare professionals [19].

Strengths and limitations

This study has several limitations. First, only the quality of the query was assessed and not the quality of the results. However, the quality of the query is strongly associated to the quality of the results. Vanopstal et al. demonstrated that under-specified queries led to an increase of noise and our data show that under-specification is the main error compensated by querying in users' first language [9]. An evaluation of the queries is planned as a second step. A sample of discrepant results (using queries built by English group versus French group) will be rated by a group of physicians. This will allow us to assess the impact of the multilingual query builder on the quality of the results. Second, this study only involves French residents in general medicine and this could affect the external validity of the results. Nevertheless, we do know that PubMed querying issues are encountered among physicians and medical researchers worldwide [14]. In order to enhance external validity, a similar trial will soon be conducted among Spanish speaking residents and physicians at the Buenos Aires Italian Hospital (Argentina).

This study is, to our knowledge, the first published evaluation of a multi-lingual query builder to access the

PubMed subset. In order to avoid any before and after studies bias, a randomised controlled trial was carried-out. The clinical questions were drafted by an experienced medical librarian (GK) and validated by two physicians (MS and NG). The theoretical difficulty levels of the clinical questions were proven due to the significant association between the average number of errors and the difficulty level of the questions. The assessment of the queries was made independently by two librarians (GK and LS), using a modified published classification [9].

Perspectives

As might be expected, this study clearly demonstrates that querying in first language is easier than querying in English. This study will soon be repeated among Argentinean healthcare professionals, comparing the use of Spanish and English. The multi-lingual query builder permits to overcome the obstacle of English when building queries, and could be of major interest for students, clinicians and researchers worldwide.

The query builder is already available in more than fifteen languages including Dutch, English, Finnish, French, German, Italian, Portuguese and Spanish. Some translations of MeSH terms or of the web site interface may still be lacking and we actively encourage all the teams working on MeSH translation, i.e. INSERM and Inist-CNRS in France. Our tool permits physicians and medical researchers to perform a request in the most relevant PubMed database fields. Full MeSH information is available, including definitions, relations and a hierarchical qualifiers list. A wide range of synonyms is also used automatically as natural language terms to complete the query, in order to maximize recall. Some features are currently lacking, like combining previous requests, however our priority was to first build an effective easy-to-use tool.

Conclusion

Physicians often feel incompetent when seeking medical information, especially when using bibliographic databases. This sentiment is sometimes associated with a feeling of illegitimacy, as these databases were not created to meet their needs. This shows that there is a gap between an idealized academic search model and the practical requirements of everyday life [20]. This multi-lingual query builder is an effective tool to improve the quality of PubMed queries and should narrow the gap, particularly for physicians and researchers whose first language is not English.

Acknowledgement

The authors are grateful to Richard Medeiros – Medical Editor Rouen University Hospital Rouen, France for editing the manuscript.

Funding

None.

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Authors contributions

MJ, GK and SJD had the idea of creating a multilingual query builder. JG developed the whole application and optimized the architecture of the system. MS, GK, LS, SJD and NG conceived the evaluation. MS collected the data. GK and LS performed the evaluation of the queries quality. MS and NG analysed the data. MS and NG drafted the manuscript. All the authors made substantial enhancement to it and approved the final manuscript. MS is guarantor.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Consent for publication

Not applicable.

Ethics approval and consent to participate

This study was approved by the ethical committee of the Rouen University Hospital (notification E2017-014). All participants gave their oral informed consent before participating.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Biomedical Informatics, Rouen University Hospital, Rue de Germont, 76031 Rouen Cedex, France. ²TIBS, LITIS EA 4108, Rouen University, Rouen, France. ³Department of General Practice, Rouen University, Rouen, France. ⁴INSERM, U1142, LIMICS, 75006 Paris, France.

Received: 14 December 2016 Accepted: 14 June 2017

Published online: 03 July 2017

References

- Lipscomb CE. Medical Subject Headings (MeSH). *Bull Med Libr Assoc.* 2000; 88(3):265–6.
- Sheets L, Gavino A, Callaghan F, Fontelo P. Do language fluency and other socioeconomic factors influence the use of PubMed and MedlinePlus? *Appl Clin Inform.* 2013;4(2):170–84.
- Liu F, Ackerman M, Fontelo P. BabelMeSH: development of a cross-language tool for MEDLINE/PubMed. *AMIA Annu Symp Proc AMIA Symp.* 2006;1012.
- Fontelo P, Liu F, Leon S, Anne A, Ackerman M. PICO Linguist and BabelMeSH: development and partial evaluation of evidence-based multilanguage search tools for MEDLINE/PubMed. *Stud Health Technol Inform.* 2007;129(Pt 1):817–21.
- Griffon N, Schuers M, Soualmia LF, Grosjean J, Kerdelhué G, Kergourlay I, et al. A search engine to access PubMed monolingual subsets: proof of concept and evaluation in French. *J Med Internet Res.* 2014;16(12), e271.
- Schuers M, Griffon N, Kerdelhue G, Foubert Q, Mercier A, Darmoni SJ. Behavior and attitudes of residents and general practitioners in searching for health information: From intention to practice. *Int J Med Inf.* 2016;89:9–14.
- Thirion B, Robu I, Darmoni SJ. Optimization of the PubMed automatic term mapping. *Stud Health Technol Inform.* 2009;150:238–42.
- Cimino JJ, Elhanan G, Zeng Q. Supporting infobuttons with terminological knowledge. *JAMIA.* 1997;4(Suppl):528–32.
- Vanopstal K, Buyschaert J, Laureys G, Vander Stichele R. Lost in PubMed. Factors influencing the success of medical information retrieval. *Expert Syst Appl.* 2013;40(10):4106–14.
- Coumou HCH, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. *J Med Libr Assoc.* 2006;94(1):55–60.
- Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. *Health Inf Libr J.* 2007;24(2):78–94.
- Bryant SL. The information needs and information seeking behaviour of family doctors. *Health Inf Libr J.* 2004;21(2):84–93.
- Meats E, Brassey J, Heneghan C, Glasziou P. Using the Turning Research Into Practice (TRIP) database: how do clinicians really search? *J Med Libr Assoc.* 2007;95(2):156–63.
- Herskovic JR, Tanaka LY, Hersh W, Bernstam EV. A day in the life of PubMed: analysis of a typical day's query log. *J Am Med Inform Assoc.* 2007;14(2):212–20.
- Vanopstal K, Stichele RV, Laureys G, Buyschaert J. PubMed searches by Dutch-speaking nursing students: The impact of language and system experience. *J Am Soc Inf Sci Technol.* 2012;63(8):1538–52.
- Magin P, Morgan S, Wearne S, Tapley A, Henderson K, Oldmeadow C, et al. GP trainees' in-consultation information-seeking: associations with human, paper and electronic sources. *Fam Pract.* 2015;32(5):525–32.
- Thiele RH, Poiré NC, Scalzo DC, Nemergut EC. Speed, accuracy, and confidence in Google, Ovid, PubMed, and UpToDate: results of a randomised trial. *Postgrad Med J.* 2010;86(1018):459–65.
- Yepes AJ, Prieur-Gaston E, Névoul A. Combining MEDLINE and publisher data to create parallel corpora for the automatic translation of biomedical text. *BMC Bioinformatics.* 2013;14:146.
- Gagnon M-P, Pluye P, Desmartis M, Car J, Pagliari C, Labrecque M, et al. A systematic review of interventions promoting clinical information retrieval technology (CIRT) adoption by healthcare professionals. *Int J Med Inf.* 2010; 79(10):669–80.
- Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. *Health Inf Libr J.* 2010;27(1):2–10.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



4.6. Conclusion

Parmi les obstacles à la recherche d'information identifiés dans l'étude qualitative figurait le fait que de nombreuses sources et supports d'information sont en langue anglaise. Le Département d'Informatique et d'Information Médicales du CHU de Rouen a donc développé plusieurs outils destinés à répondre à ces difficultés :

- LiSSa (Littérature Scientifique en Santé), base de données bibliographiques en santé francophone, regroupe à la fois des références issues de la base de données PubMed et des références issues d'éditeurs comme Elsevier, non indexées dans la base de données états-unienne.
- Le constructeur de requêtes bibliographiques médicales (CRBM). Il s'agit du premier outil disponible permettant d'élaborer des requêtes complexes en français dans la base de données PubMed. Notre travail démontre qu'il s'agit d'un outil efficace pour améliorer la qualité des requêtes effectuées dans PubMed par des utilisateurs dont le français est la langue maternelle.

Il nous a cependant semblé possible que les difficultés d'accès à la connaissance des médecins généralistes soient liées au moins en partie à la difficulté d'identifier précisément les informations utiles à leur activité clinique. Nous nous sommes donc, dans un dernier temps, intéressés à la production scientifique en médecine générale. Quels sont les contours de la production scientifique en médecine générale ? Quelles sont les thématiques de recherche dans cette discipline ? Ce sont les questions évoquées dans la dernière partie de notre travail.

5. Thèmes et tendances dans la production scientifique en médecine générale

5.1. Introduction

5.1.1. La recherche en médecine générale

Au cours de la dernière décennie du 20^{ème} siècle, nous avons assisté à une reconnaissance accrue par l'Organisation Mondiale de la Santé de l'importance des soins primaires dans l'amélioration de la santé des populations (95). Cette reconnaissance institutionnelle faisait suite à la parution de plusieurs travaux concernant l'impact des soins primaires dans l'efficience des systèmes de soins. Le travail de Barbara Starfield notamment, reliant la force du secteur des soins primaires dans de nombreux pays à l'amélioration de l'état de santé des populations, a renforcé l'idée d'une nécessité de la recherche en médecine générale (96).

En 1961, un article intitulé "The Ecology of Medical Care" estimait, pour les Etats-Unis et la Grande Bretagne, que dans une population de 1 000 habitants exposés à un problème de santé, 750 signalaient des troubles, 250 parmi eux consultaient un médecin, 9 étaient hospitalisés, 5 dirigés vers un autre médecin et 1 hospitalisé dans un centre universitaire (97). Ces données ont été réactualisées en 2001 et les auteurs ont trouvé des résultats très similaires (98). Ces résultats ont conduit à préconiser un rééquilibrage de la formation médicale et de la recherche clinique de l'hôpital et les spécialités vers la médecine de première ligne. En effet, il ne paraissait pas raisonnable de laisser en friche l'étude des problèmes de santé les plus fréquents dans la population.

De plus, la médecine générale occupe dans de nombreux systèmes de santé, dont le nôtre, une place particulière de première ligne, qui justifie l'existence d'un questionnement scientifique spécifique, non pris en charge par les autres spécialités, à l'exception partielle de la santé publique (99).

Sous l'influence de cette prise de conscience, de nombreux pays développés et en voie de développement ont renforcé le rôle des soins primaires dans l'organisation et la structure de leur système de santé et ont adapté le programme éducatif de leurs universités. Cependant, cet intérêt croissant pour la fourniture de soins de santé primaire et pour l'éducation n'a pas été systématiquement accompagné par l'investissement dans les efforts de recherche. Bien qu'elle soit un grand potentiel de l'amélioration de la santé de la population, la recherche en soins primaires apparaît comme le chaînon manquant dans le développement d'une médecine générale de haute qualité fondée sur des preuves (100).

Diverses actions ont été menées pour tenter de mieux exploiter ce potentiel. Parmi elles, on peut citer la conférence organisée par l'Organisation Mondiale des Médecins de Famille (WONCA) en 2003 intitulée : « Améliorer la santé mondiale : la nécessité de la recherche en médecine générale ». Il semble que des progrès constants aient été réalisés. En effet, alors qu'une étude concluait en 2004 que les publications liées à la médecine générale étaient "presque invisibles" dans PubMed par rapport aux publications d'autres disciplines (101), la bibliothèque étatsunienne de médecine (*National Library of Medicine*, NLM) regroupait en 2009 les revues de soins primaires qui étaient auparavant inclus dans d'autres catégories, au sein d'une catégorie "soins de santé primaires". Cette action a été suivie en 2011 par la base de données bibliographiques Web of Science de Thomson Reuters qui a introduit une catégorie « soins de santé primaires » pour y inclure 14 revues disséminées auparavant dans d'autres disciplines (102).

En France, en 2006, un rapport a été remis au Ministère de la Santé, décrivant des propositions pour développer la recherche en médecine générale et en soins primaires (99). Il rappelle la nécessité de produire des connaissances scientifiques utiles aux professionnels de santé qui sont en première ligne pour répondre aux besoins des patients, et dont l'activité recouvre une part majeure des soins dispensés dans le système de santé. Ce rapport précise que cette recherche sera d'abord clinique et épidémiologique, à l'instar de celle qui est conduite à l'hôpital et par les spécialistes de santé publique, mais qu'elle devra aussi intégrer les dimensions psychologique et sociale de la demande de soins. L'analyse des comportements des médecins et des malades tiendra dans ces travaux une part importante. Elle devra par ailleurs s'inscrire dans une

perspective plus large, qui est celle de la recherche en soins primaires ce qui implique la création d'équipes multidisciplinaires.

5.1.2. La bibliométrie

La bibliométrie a été définie comme l'utilisation de méthodes statistiques pour analyser un corpus littéraire dans le but de révéler le développement historique à travers l'étude scientifique et quantitative des publications (103).

Le développement d'outils et de méthodes est devenu indispensable pour mieux appréhender la complexité croissante de la production scientifique.

Les principaux indicateurs bibliométriques sont :

- Le nombre de publications : les rapports d'évaluation des unités de recherche exigent souvent le classement et le comptage des publications (103).
- Le nombre de citations est un indicateur commun dans les "sciences dures". Une publication, une fois qu'elle est publiée, va recevoir x citations, c'est à dire qu'elle sera intégrée à la bibliographie finale de x autres publications.
- Le facteur d'impact (FI) d'une revue mesure la fréquence avec laquelle ses articles sont cités dans les articles d'autres revues pendant une année donnée. Le facteur d'impact pour l'année N est calculé en divisant le nombre de citations (pendant l'année N) de ses articles publiés les années N-1 et N-2 par le nombre d'articles publiés dans cette revue les années N-1 et N-2 (103).
- Les quartiles : une revue pouvant être rattachée à plusieurs domaines, son facteur d'impact doit être évalué en comparaison avec les facteurs d'impact des revues du même domaine (103). C'est à partir de ce concept qu'a été développé l'outil SIGAPS (Système d'Interrogation, de Gestion et d'Analyse des Publications Scientifiques).
- L'indice d'immédiateté qui calcule le rapport entre le nombre de citations obtenues par la revue l'année N sur le nombre d'articles publiés par cette revue l'année N. En d'autres termes, il peut indiquer si les articles de la revue sont très vite cités ou non (103).

- L'âge médian des articles cités pendant l'année. Il informe du nombre d'années au bout duquel le nombre cumulé de citations pour les articles de la revue atteignent 50% du nombre total de citations reçues par les articles de la revue (103).
- L'indice h (ou indice de Hirsch ou index h , h index en anglais) vise à quantifier la productivité scientifique et l'impact d'un scientifique, d'une unité, en fonction du niveau de citation de ses publications (103). Pour le déterminer on classe les publications par ordre décroissant de citations. Un indice h de n signifie qu'un auteur a au moins n publications citées n fois.

Les indicateurs bibliométriques ont bien sur des limites (104) et soulèvent donc de nombreuses interrogations lorsqu'ils sont mis au service des politiques de la recherche publique (105). La bibliométrie quantitative, introduite dans le processus de prise de décision de la politique de recherche, est à double tranchant puisque comme toute donnée statistique, ces indicateurs peuvent être interprétés de multiples façons s'ils sont sortis de leurs contextes. Ainsi, selon H. Rostaing, « il ne faut absolument pas considérer les données bibliométriques comme des substituts à l'avis d'expert, mais tout au contraire, comme des outils de travaux dans l'élaboration du consensus parmi des experts » (105).

5.1.3. Analyse bibliométrique de la production scientifique en médecine générale

L'évaluation de la production scientifique dans le domaine de la médecine générale soulève plusieurs difficultés.

D'une part, la recherche exhaustive des publications dans le champ de la médecine générale dans la base de données PubMed est difficile car la médecine générale est une spécialité médicale concernée par plusieurs mots clés, qualificatifs ou types de publications. De plus, des problèmes de nature linguistique attribuables à des usages différents de la terminologie et aux distinctions dans l'orthographe de l'anglais américain et britannique exposent à des divergences dans les résultats des recherches concernant la médecine générale. En effet, l'identification des productions scientifiques sur certains sujets se traduisent par des résultats très différents selon leur association aux mots-clés anglais « general practice », « family practice » et « family medicine » (106). Afin de pallier

ce problème, le D2IM du CHU de Rouen définissait dès 1999 des super concepts (ou métatermes) qui regroupent sémantiquement différents mots clés, qualificatifs ou types de publications d'une même discipline médicale (107).

La dispersion importante des articles liés à la médecine générale au sein de multiples revues de médecine générale et encore plus largement au sein de revues qui ne sont pas spécialisées dans ce champ accroît encore la difficulté de la recherche. En 2012, les 250 articles les plus cités en médecine générale ont été publiés parmi 71 revues différentes dont seulement 5 pouvant être étiquetées comme revue de médecine générale (108).

Ainsi, peu d'études bibliométriques ont été réalisées dans le domaine de la médecine générale. Bien que certaines analyses aient été effectuées sur des pays spécifiques (109)(110), une période donnée (111), ou la production d'une revue spécifique (112), à notre connaissance, au cours des dix dernières années, aucune tentative n'a été faite pour décrire l'ensemble de ce champ. C'est ce que nous nous sommes proposés de faire dans l'article présenté ci-après.

5.2. Tendances et thématiques en médecine générale dans PubMed

5.2.1. Contexte

La recherche est indispensable au développement d'une médecine générale de haute qualité fondée sur des preuves. Un nombre croissant de médecins généralistes accède à des bases de données électroniques pour leurs recherches documentaires et mettre à jour leurs connaissances. Ces bases de données électroniques permettent une diffusion plus large de la littérature scientifique. Cependant, les publications en médecine générale sont peu nombreuses, largement dispersées et les internes en médecine générale et médecins éprouvent des difficultés à identifier une information de qualité, objective et fiable sur internet.

5.2.2. Objectif

L'objectif de cette étude était de fournir une analyse de la production scientifique liée à la médecine générale en suivant l'évolution du nombre de publications, et en identifiant les principaux sujets ainsi que les acteurs prédominants (journaux, pays) impliqués dans ce champ.

5.2.3. Méthode

Une recherche a été réalisée sur PubMed pour extraire les articles dans le champ de la médecine générale via une requête spécifique élaborée par des documentalistes et des médecins généralistes. Plusieurs métadonnées ont été sélectionnées dans chaque article : le nom du journal, les mots-clés MeSH, la langue de publication et l'année de publication. D'autres données ont été recueillies parallèlement : le facteur d'impact des journaux, via le Journal Citation Reports, le pays d'affiliation du premier auteur, la taille de la population des pays, et le PIB par habitant des pays.

Le nombre d'articles dans le champ de la médecine générale publiés par année a été calculé. Le taux de croissance de la littérature concernant la médecine générale a été calculé par périodes de 10 ans. Le pays du premier auteur a été détecté à partir de son affiliation. La fréquence des mots-clés MeSH utilisés par les indexeurs PubMed pour

classer les articles a été calculée à partir des articles extraits. La langue de publication a été déterminée en utilisant le champ « Language » de PubMed pour tous les articles récupérés.

5.2.4. Résultats

75 093 articles ont été extraits, publiés dans 3 802 journaux. Les 3 journaux les plus productifs étaient « British Journal of General Practice » (6,8%), « British Medical Journal » (6,7%) et « Family Medicine » (3,4%). Les articles étaient principalement écrits en anglais (83,3%), le français arrivant en 3^{ème} position (2,9%).

Les Etats-Unis étaient le pays le plus productif en nombre d'articles publiés (27,2%), la France se trouvait au 7^{ème} rang (2,5%).

Le nombre d'articles publiés dans le champ de la médecine générale progressait au cours des années, passant de 746 en 1985 à plus de 2 500 publications en 2014. Les articles français suivaient la même évolution avec un record de 72 articles publiés en 2013.

Les 2 grands thèmes les plus représentés étaient la pratique médicale (« modèles de pratique médicale », « attitude du personnel soignant ») et la formation des médecins (« internat et résidence », « formation médicale continue »).

Une présentation détaillée de la méthode et des résultats de cette étude figure ci-après.

5.3. Trends and topics in general practice in PubMed

Auteurs

Clémence SIEDLECKI, Nicolas GRIFFON, Gaëtan KERDELHUE, Julien GROSJEAN, Benoît THIRION, Marc JAMOULLE, Stéfan J DARMONI, Matthieu SCHUERS

Abstract

Objective: The purpose of this study is to provide an analysis on scientific production in order to identify the major topics as well as the predominant actors (journals, countries) involved in the field of general practice.

Methods: A PubMed search was carried out to extract article citations related to general practice using a specific query. Data were downloaded from PubMed in Extensible Markup Language (XML) and was processed through a dedicated parser.

Results: A total of 75,093 articles were retrieved among 3,802 journals. Half of journals published only one or two articles related to general practice. Articles were written mainly in English (83.3%). A total of 733,615 MeSH terms were identified, including 10,540 unique MeSH terms, mainly related to care organization and medical training. The USA were the most productive country in number of articles published (27.2%) and related to GDP per capita. Related to population size, ranking positions changed in favor of smaller countries (Denmark, Norway and the Netherlands).

Conclusion: Due to the large number of articles included and the numerous parameters analyzed, this study provides a wide view of scientific productivity related to general practice and allows us to better understand this field.

Keywords

Bibliometrics, general practice, PubMed

Introduction

Evidence based medicine has become the new paradigm in general practice and it is accepted that research is essential to the development of a high quality general practice

based on evidence [1]. To fill the gap between research and clinical practice, general practice must increase its knowledge base [2]. In order to meet this requirement and more widely to answer the question of the importance of research in general practice, World Organization of Family Doctors (WONCA) held a conference in 2003 in order to promote research. Recommendations were developed [3] such as the development of conferences and of research journals in general practice with the improvement of the representation of these journals in bibliographic databases; the development of national institutes of research and of university departments of general practice or even the facilitation of funding for the international collaborative research in general practice; and progress has been made [4]. In 2009, the National Library of Medicine (NLM) consolidated in its bibliographic data base Medline, primary care journals previously scattered in other categories within the category “primary health care”, including “family practice” topics. In 2011, the bibliographic data base Web of Science by Thomson Reuters also introduced a category “primary health care” to include 14 journals previously scattered in other disciplines [5].

An increasing number of general practitioners now access electronic databases for their documentary research and to update their knowledge [6]. Nevertheless, despite the growing availability and use of online information sources, some clinical questions remain unanswered [7]. Physicians and residents in general practice are still struggling to find good quality, objective and reliable information on the internet [8].

Bibliometrics is to scientific papers what epidemiology is to patients. By using the metadata of scientific publications, it enables the measurement of performance of research teams, institutes and countries. It also enables the identification of national and international networks, the description of the dynamics of production and the mapping of new scientific and technological areas of development [9].

The purpose of this study was to provide an analysis of the scientific production associated with general practice by following the evolution of the number of publications and by identifying the main subjects as well as the main players (journals, countries) involved in this field.

Material and methods

We performed a bibliometric analysis on general practice related publications of PubMed. The PubMed database (URL: <http://www.pubmed.org>) is developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM).

Bibliographic search

A PubMed search was carried out in June 2016. Two medical librarians (GK, BT) and four physicians (CS, MJ, SJD, MS) agreed on a comprehensive list of MeSH terms to represent the general practice domain. This list included the following MeSH terms: "community medicine", "general practice", "general practitioners", "family practice", "family physicians", "gatekeeping" and "primary care physicians".

In order to optimize the precision (i.e. minimizing the false positive rate), we decided to select only the terms indexed as "Major Topic". In order to optimize the recall (i.e. minimizing the false negative rate) and to select recent publications not yet indexed by NLM librarians, we decided to search all the selected MeSH terms synonyms in the publications titles [10].

The following query was built: "community medicine"[MAJR] OR "community medicine"[TI] OR "medicine, community"[TI] OR "family doctor "[TI] OR "primary care practitioner "[TI] OR "primary care physician"[TI] OR "family medicine"[TI] OR "GP/FM"[TI] OR "general practitioner"[TI] OR "physicians, family"[MAJR] OR "physicians, family"[TI] OR "family physician"[TI] OR "physician, family"[TI] OR "family physicians"[TI] OR "family doctor "[TI] OR "primary care practitioner "[TI] OR "primary care physician"[TI] OR "gatekeeping"[MAJR] OR "gatekeeping"[TI] OR "general practitioners"[MAJR] OR "general practitioners"[TI] OR "practitioner, general"[TI] OR "practitioners, general"[TI] OR "general practice physicians"[TI] OR "general practice physician"[TI] OR "general practitioner"[TI] OR "general practice"[MAJR] OR "general practice"[TI] OR "practice, general"[TI] OR "physicians, primary care"[MAJR] OR "primary care physicians"[TI] OR "primary care physician"[TI] OR "family practice"[MAJR] OR "family practice"[TI] OR "practices, family"[TI] OR "family practices"[TI] OR "practices, general"[TI] OR "general practices"[TI] OR "practice, family"[TI].

Analysis of growth of literature

The average decade growth rate of the literature related to general practice was calculated using the following equation:

Growth rate (%) = ((Current Decade Total Number of Articles – Previous Decade Total Number of Articles)/Previous Decade Total Number of Articles) x 100.

These growth rates were also calculated for the whole PubMed data base, including the articles identified by our query.

Country detection

The country of the first author was automatically detected from its affiliation. A specific algorithm in two steps was developed: (1) From email address country code top-level domain based on ISO 3166-1 (e.g. .it refers to Italy); (2) from each address field, a comparison was performed with a dedicated geographical database enriched by the Google map API (geocode).

Rankings have been realized to classify country by number of articles published during the period, and after adjustment for population size (number of articles published divided by population size) [11] and GDP per capita (number of articles published divided by GDP per capita) [12].

Bradford's law

Bradford's law is used in the information science literature to describe the dispersion of articles in any scientific field, or to identify "core journals" of serial titles [13]. Bradford's law states that "If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus". To identify the core journals containing articles dedicated to general practice, Bradford's law was applied by dividing the journals ranked according to publication output frequency into three groups with each group representing approximately the same number of articles.

The 2013 Journal Citation Reports (JCR) (Thomson Reuters, New York, USA) was used for Impact Factor (IF) determination [14].

Analysis of MeSH terms and language of publication

The frequency of MeSH keywords used by the PubMed librarians to classify the articles has been calculated from the extracted articles. Publication language was determined using the PubMed "Language" field for all articles retrieved.

Data extraction

Data were downloaded from PubMed in Extensible Markup Language (XML) and was processed through a dedicated parser. The selected metadata were integrated in a database.

Results

The search in PubMed resulted in a total of 75,093 references published in 3,802 different journals over the period 1841 to June 2016.

Growth of literature

The number of publications in the general practice field has increased from 793 annual publications in 1980 to more than 2,000 in 2010 (figure 1). This increase accelerated in the 1990s, exceeding the 2,500 articles published in 2014. This increasing growth in publications has followed the overall trend of the whole scientific production indexed in PubMed which has grown by more than 193% between 1980 and 2010 (Figure 2). Between the 1980s and the 2000s, growth in general practice was more important than the overall growth, reaching a record of 93% in the 1990s (Figure 3).

Languages of publication

Among the 75,093 articles, 33 different languages were identified. The five most represented languages were English (n= 62,532; 83.27%), German (n= 4,130; 5.50%), French (n= 2,192; 2.92%) Spanish (n= 1,188; 1.58%) and Dutch (n= 978; 1.30%). The distribution of articles according to the language of publication is shown in Figure 4.

Geographical distribution and socioeconomic factors

The identification of the country of the first author was possible for 34,732 articles (46.25%), 116 different countries were identified. Table 1 shows the top 20 countries according to the number of publications of general practice. The 3 main producers of articles in this field were the USA, the United Kingdom and Australia, accounting for over half of the total (57.6%). The scientific production has been classified depending on the population size of the country (Table 2) and the gross domestic product (Table 3).

Developed countries such as Denmark, Norway and the Netherlands had better rates of publication according to the size of their population. According to the GDP per capita, the

United States, the United Kingdom and India were the biggest producers of articles.

Analysis of journals

3,802 different journals were identified. The 3 journals publishing the largest number of articles in the general practice area were the “British Journal of General Practice” (n= 5,073, 6.8%), the “British Medical Journal” (n= 5,029, 6.7%) et “Family Medicine” (n= 2,574, 3.4%). Almost half of journals (50.55%) published only 1 or 2 articles related to general practice.

According to Bradford’s law, the total number of articles compared to the number of journals in order of decreasing productivity was divided into 3 groups containing approximatively the same number of articles. The first third of the total number of articles (n= 25,133 articles) represented the journals (n= 10; 0.26%) that published between 5,073 and 1,300 articles (Table 4). The middle third of articles (n= 25,046 articles) corresponded to the journals (n= 105; 2.76%) that contained an average number of articles (between 1,119 and 84 articles), and the last third of articles (n= 24,914 articles) included the “long tail” of journals (n= 3,687; 96.97%) that published the fewest articles (less than 82). Within the first group, 2 journals (“British Medical Journal” and “Medical Journal of Australia”) were not journals of general practice according to the JCR and the NLM catalogue [14].

The percentage of articles in the general practice area compared to the total number of articles produced by the journals ranged from 2.8% (“British Medical Journal”) to 63.7% (“Family Medicine”) (Table 4). The three most specific journals in general practice, ie. publishing the most important proportions of articles in this area, were “Family Medicine”, “Family Practice” and the “British Journal of General Practice” (Figure 5).

MeSH terms

The search results contained 733,615 MeSH terms, including 10,540 unique MeSH terms. Consequently, the average number of MeSH terms per article was 9.8. Within the 20 most heavily quoted MeSH terms which did not appear in the request, only two keywords related to health conditions: “mental disorders” (n= 1 157; 1.54%) and “high blood pressure” (n= 919; 1.22%) (Table 5). The two most represented themes were medical practice and the training of doctors.

Discussion

Summary of main findings

Our request has extracted 75,093 articles within 3,802 different journals. The number of publications in the general practice field has increased, from 793 yearly publications in 1980 to more than 2,000 in 2010, with a record of more than 2,500 published articles in 2016.

The vast majority of the articles were written in English (83.3%). The 3 biggest producers of articles were the United States (27.2%), The United Kingdom (21.02%) and Australia (9.4%); France ranked 7th (2.5%).

The 3 most productive journals in absolute value were the “British Journal of General Practice” (6.8%), the “British Medical Journal” (6.7%) and “Family Medicine” (3.4%). More than half of the journals (50.55%) published only one or two articles in the general practice field. The 3 most specific journals, i.e. publishing the largest proportions of articles in the general practice field, were “Family Medicine”, “Family Practice” and the “British Journal of General Practice”.

Within the 20 most heavily quoted MeSH terms, only two keywords were related to health conditions. The two most represented topics were medical practice and the training of practitioners.

Comparison with existing literature

All the 75,093 extracted articles represented only 0.29% of the total of the articles on PubMed, which is low compared to other disciplines such as for example ophthalmology which reaches 1.7% [15]. The level of production in research in general practice seems to be less important than other medical disciplines, like surgery or public health [16]. These results can be explained by our choice to build a specific query, in order to optimize the precision. Other bibliometric analyses were made in the field of general practice using more sensitive queries [17]. This kind of queries increase the number of citations but increase the false-positive rate as well. However, it remains difficult to clearly identify the references in the field of general practice indexed in PubMed. Unlike most medical disciplines, general practice is not defined by an organ or a condition. Several authors attempted to build a relevant general practice search filter for PubMed but they all failed

to combine high levels of precision and sensitivity [18-20].

However, our study shows a continuing growth of the number of publications in the field of general practice for some thirty years. In 2006, upon the request of the Wonca's European branch, the European General Practice Research Network (EGPRN) created a group of experts to work on the development of a European research agenda in general practice [21]. The main purposes of this project were to determine future directions for research and research policies, also taking into account the challenges, the approach and the future developments of this discipline. The integration of a "primary health care" category within the Medline bibliographic databases in 2009, and then in Web of Science in 2011 contributed to the development of the exposure of general practice journals. In France, profound changes happened in order to continually structure research carried out by general practitioners on their practices and their role in the care system. The creation in 2004 of the specialty of general practice and the creation of the first positions of research fellowship in general practice in 2007 were designed to offer training in research methods and a dedicated environment for the development of high quality research [22]. The clear majority of the articles in the general practice field was written in English (83.3%), as all the indexed articles in PubMed (82.6%). This predominance of English has undoubtedly some consequences. The level of language proficiency in English is one of several factors influencing practitioners' ability to build queries for relevant searches on PubMed [23]. Some tools already assist non-English speakers to access to the PubMed references written in their mother tongue: BabelMeSH [24], PICO Linguist [24] and LiSSa [24], but by definition they only give access to a restricted area of the existing literature. Other tools allow an automatic translation of biomedical text, including titles and abstracts and using the MEDLINE database [27]. More recently, we developed a multilingual query builder which enables the creation of sophisticated bibliographic queries by combining medical MeSH terms; it could be very helpful to clinical research and practice for non-English speakers [28].

The United Kingdom and the Netherlands were among the most productive countries according to the number of published articles but also according to the size of their population and their GDP per capita. This can be explained by past political choices in favor of general practice research. In the United Kingdom, the tradition of an epidemiologic and health care research in the first line health care asserts itself since 1952 with the foundation of the Royal College of General Practitioners, and continued in 1963

with the first university chair of general practice [29]. In the 2000s, an ambitious development plan was executed by the ministry of health for the investment in the training of research professionals, the investment in research infrastructures and research funding [29]. In addition, since 1987, the information of several million British patients have been gathered from hundreds of general practitioners' offices and then anonymized to constitute the General Practice Research Database (GPRD), which became the world's biggest database between 1994 and 2012 [30]. The Clinical Practice Research Datalink (CPRD), a public organization created in 2012 (formed through the merger of the GPRD and the Health Research Support Service (HRSS)), has simplified the provision of health data from both public and private research organizations, giving rise to more than 1500 publications [31]. In the Netherlands, the creation of a university stream in general practice goes back to the 1960s with a 3-year specialty university training course imposed on general practitioners. As in the United Kingdom, general practice departments could rely on networks of practitioners' offices in town, by conducting systematic recordings of consultations, to carry out epidemiological work in primary care [32].

The articles in the general practice field were scattered within numerous journals of general practice and more broadly within journals which were not specialized in this field. Half of the identified journals published only 1 or 2 articles about general practice. In addition, among the 3 802 journals, only 19 could be considered as journals of general practice according to the JCR [14] and 46 according to the NLM catalogue. A study carried out in 2012 also found an important dispersion of the literature in general practice [33]. This shows that many references regarding the field of general practice are published in journals of other disciplines. This can be explained by the involvement of general practitioners in research projects within other medical disciplines. In addition, non-general practice journals often have a better impact factor than general practice journals [34]. A study carried out in the United States showed that the general practitioners published in journals which do not fall into general practice as they believed that these journals were more prestigious and reached a wider audience. Moreover, the promotion and institutional tenure committees tend to give less importance to the publications of general practice journals [35].

Within each journal, the percentage of articles related to general practice compared to the total number of produced articles ranging from 2.8% (British Medical Journal) to 63.7% (Family Medicine). Among the 10 journals publishing the biggest number of articles in the

field of general practice, only 3 had a proportion of articles in this field exceeding 40% (FM, FP and BJGP) and only one exceeded 50%. These figures were well below some specialties such as ophthalmology. A recent study showed that, except for one journal, the percentage of articles relating to eye diseases compared to the number of produced articles ranged from 58.1% to 91.9% depending on the journals, among the “core journals” of ophthalmology according to Bradford’s law [15]. The transverse nature of general practice, as well as the scope of clinical situations experienced by the general practitioners may partly explain by these data. Another explanation could come from the indexation procedure of the articles. Indeed, the articles regarding general practice may have been indexed without any keyword in this field, potentially resulting in greatly underestimated articles recovered. This under-indexation may be related to a difficulty in qualifying an article considered as belonging to the field of general practice by the lack of knowledge of this specialty and its field of expertise.

The analysis of the MeSH terms frequency showed that the 2 most popular topics were those regarding general practice and practitioners’ training. Although the importance of patient-centered care is widely recognized, this MeSH terms distribution, focused on the practitioner and his practice, could show the youth of this research, still hardly orientated towards the patient.

Strengths and limits

Our study has some limitations. PubMed is the database the most widely used for the bibliometric analysis, but it does not hold all the journals of the biomedical field.

By excluding in our request the MeSH terms related to primary care, we may have underestimated the number of articles found by removing some articles related to general practice. However, this term is highly associated with articles dealing with the organization of health care, and not health care by itself. Similarly, it is very likely that some articles in the field of general practice are not indexed with keywords related to general practice.

The greatest strength of our study results from the precision of our query, while we optimized the recall thanks to the addition of synonyms. PubMed was selected for two reasons. Firstly, we chose it due to its wide coverage, its international visibility, its quality criteria, its free access and because it is a premium bibliographic database for the quantitative assessment of the scientific publications in the field of general practice [36].

Secondly, the articles on PubMed are indexed with the MeSH, reference thesaurus in the biomedical field, which includes a set of standard words which can reflect the content of the articles [37]. With the MeSH terms, PubMed provides a good possibility to extract emerging keywords.

To our knowledge, this is the first bibliometric study in the field of general practice carried out on all the indexed publications on PubMed.

Perspectives

This study provides a broad overview of the scientific production linked to general practice and results in an improved knowledge of this specialty. The identification of the most specific journals will contribute to facilitating the choice of journals in which general practitioners could find information and update their knowledge. The dispersion of articles throughout many journals and the difficulties of indexation of the articles in the field of general practice make the identification of the relevant articles difficult by the practitioners [8]. The development of specific research strategies in general practice, such as search filters, is one of the solutions to facilitate the efficient identification of the relevant publications [19]. A relevant general practice search filter, both specific and sensitive, has yet to be developed. Our study shows an important growth in the publications in general practice, likely due to an international awareness on the importance of research in this young specialty. Our data could suggest that this field remains under-represented in PubMed, but this probably due to a difficult identification of general practice references. The analysis of research policies of the most productive countries in general practice scientific literature could be a first step to develop for the discovery of ways of optimization for the less productive countries. Finally, the analysis of the main topics of publication is a first step toward discovering new ways of research, more patient-centered.

References

1. Rosser WW, van Weel C. Research in family/general practice is essential for improving health globally. *Ann Fam Med*. 2004 May 26;2 Suppl 2:S2-4.
2. De Maeseneer JM, van Driel ML, Green LA, van Weel C. The need for research in

- primary care. *Lancet*. 2003 Oct 18;362(9392):1314-9.
3. Van Weel C, Rosser WW. Improving health care globally: a critical review of the necessity of family medicine research and recommendations to build research capacity. *Ann Fam Med*. 2004 May 26;2 Suppl 2:S5-16.
 4. Pimlott N, Upshur RE. From clinical observation to clinical discovery: the challenge for family medicine research. *Can Fam Physician*. 2014 Jan;60(1):14-6.
 5. Van Weel C. The impact of research in primary care and family medicine: the Thomson Reuters Web of Science Subject Category 'Primary Health Care'. *Fam Pract*. 2011 Jun;28(3):239-40.
 6. Bernard E, Arnould M, Saint-Lary O, Duhot O, Hebbrecht G. Internet use for information-seeking in general practice: a cross-sectional survey among French general practitioners. *Int J Med Inf* 2012, 81:493-9.
 7. Ramos K, Linscheid R, Schafer S. Real-time information-seeking behavior of residency physicians. *Fam Med* 2003, 35:257-60.
 8. Schuers M, Griffon N, Kerdelhue G, Foubert Q, Mercier A, Darmoni SJ. Behavior and attitudes of residents and general practitioners in searching for health information: From intention to practice. *Int J Med Inform*. 2016 May;89:9-14.
 9. Lewison G, Devey M. Bibliometric methods for the evaluation of arthritis research. *Rheumatology (Oxford)* 1999;38(1):13-20.
 10. Thirion B, Robu I, Darmoni SJ. Optimization of the PubMed Automatic Term Mapping. *Stud Health Technol Inform* 2009, 150:238-42.
 11. World population review-Country populations in 2016 [Internet]. [Cited June 2016]. Available on: <http://worldpopulationreview.com/countries/>
 12. The World Bank - GDP per capita in 2014 [Internet]. [Cited June 2016]. Available on: http://data.worldbank.org/indicator/NY.GDP.PCAP.CD?order=wbapi_data_value_2014
 13. Rostaing H, La bibliométrie et ses techniques, Sciences de la société – CRRM, supplément au n°387, 1996.
 14. Journal titles ranked by impact factor in 2014 [Internet]. [Cited June 2016]. Available on: <https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action>
 15. Boudry C, Denion E, Mortemousque B, Mouriaux F. Trends and topics in eye disease research in PubMed from 2010 to 2014. *Peer J*. 2016; 4:e1557.

16. Askew DA, Glasziou PP, Del Mar CB. Research output of Australian general practice: a comparison with medicine, surgery and public health. *Med J Aust.* 2001 Jul;175(2):77-80.
17. Mendis K, Solangaarachchi I. PubMed perspective of family medicine research: where does it stand? *Fam Pract.* 2005 Oct;22(5):570-5.
18. Brown L, Carne A, Bywood P, McIntyre E, Damarell R, Lawrence M, et al. Facilitating access to evidence: Primary Health Care Search Filter. *Health Inf Libr J.* 2014 ;31(4):293–302.
19. Gill PJ, Roberts NW, Wang KY, Heneghan C. Development of a search filter for identifying studies completed in primary care. *Fam Pract.* 2014 1;31(6):739–45.
20. Sladek R, Tieman J, Fazekas BS, Abernethy AP, Currow DC. Development of a subject search filter to find information relevant to palliative care in the general medical literature. *J Med Libr Assoc.* 2006 94(4):394–401.
21. Research Agenda for General Practice. [Internet]. [Cited June 2016]. Available on: http://www.egprn.org/files/userfiles/file/research_agenda_for_general_practice_family_medicine.pdf
22. Décret n° 2008-744 du 28 juillet 2008 portant dispositions relatives aux personnels enseignants des universités, titulaires et non titulaires de médecine générale. Available on: <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000019267080>
23. Vanopstal K, Buysschaert J, Laureys G, Vander Stichele R. Lost in PubMed. Factors influencing the success of medical information retrieval. *Expert Syst Appl.* 2013;40(10):4106
24. Liu F, Ackerman M, Fontelo P. BabelMeSH: development of a cross-language tool for MEDLINE/PubMed. *AMIA Annu Symp Proc.* 2006:1012.
25. Fontelo P, Liu F, Leon S, Anne A, Ackerman M. PICO Linguist and BabelMeSH: development and partial evaluation of evidence-based multilanguage search tools for MEDLINE/PubMed. *Stud Health Technol Inform.* 2007;129(Pt 1):817-21.
26. Griffon N, Schuers M, Soualmia LF, Grosjean J, Kerdelhué G, Kergourlay I, et al. A search engine to access PubMed monolingual subsets: proof of concept and evaluation in French. *J Med Internet Res.* 2014;16(12):e271.
27. Yepes AJ, Prieur-Gaston E, Névéol A. Combining MEDLINE and publisher data to create parallel corpora for the automatic translation of biomedical text. *BMC*

Bioinformatics. 2013; 14:146.

28. Schuurs M, Joulakian M, Kerdelhué G, Segas L, Grosjean J, Darmoni SJ et al. Lost in translation? A multilingual Query Builder improves the quality of PubMed queries: a randomised controlled trial. *BMC Med Inform Decis Mak*. 2017 July;17(1):94.
29. De Pourville G, Développer la recherche en médecine générale et en soins primaires en France : Propositions, Rapport remis à Monsieur le Ministre de la Santé et Monsieur le Ministre Délégué à la Recherche, 31 mai 2006.
30. The UK General Practice Research Database, A powerful e-Clinical Research Tool. *Clinical Research Focus* (2002), 13, 48-50.
31. The Clinical Practice Research Datalink. [Internet]. [Cited June 2016]. Available on: <https://www.cprd.com/intro.asp>
32. Van Weel C. Longitudinal research and data collection in primary care. *Ann Fam Med* 2005; 3(Suppl 1):S46-S51.
33. Dunikowski LG, Freeman TR. Impact of family medicine research: Bibliometrics and beyond. *Can Fam Physician*. 2016 Mar;62(3):266-8.
34. Peleg R, Shvartzman P. Where should family medicine papers be published - following the impact factor? *J Am Board Fam Med*. 2006 Nov-Dec;19(6):633-6.
35. Weiss BD. Publications by family physicians in non-family medicine journals. *Fam Pract Res J*. 1990 Winter;10(2):117-22.
36. Jelercic S, Lingard H, Spiegel W, Pichlhöfer O, Maier M. Assessment of publication output in the field of general practice and family medicine and by general practitioners and general practice institutions. *Fam Pract*. 2010 Oct;27(5):582-9.
37. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis. *J Med Internet Res*. 2015 Mar 25;17(3):e81.

Figure 1: Evolution of publications in the general practice field by year

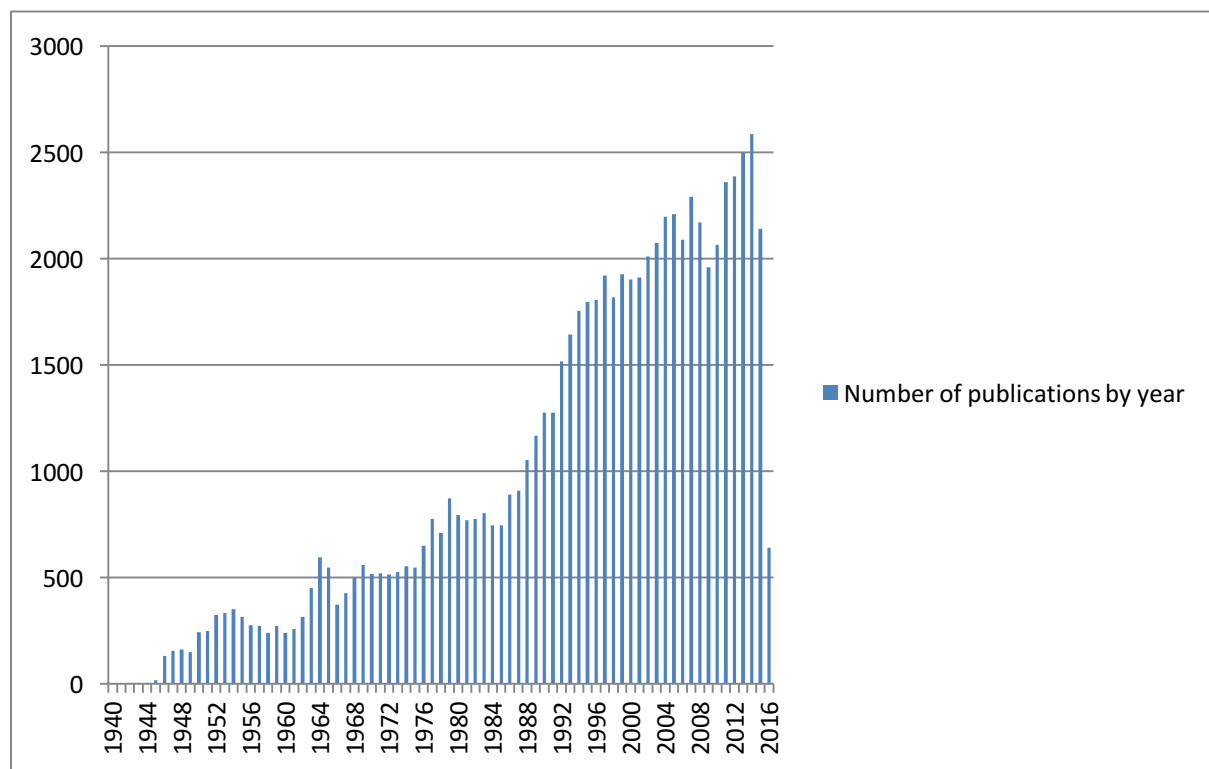


Figure 2: Evolution of publications in PubMed by year

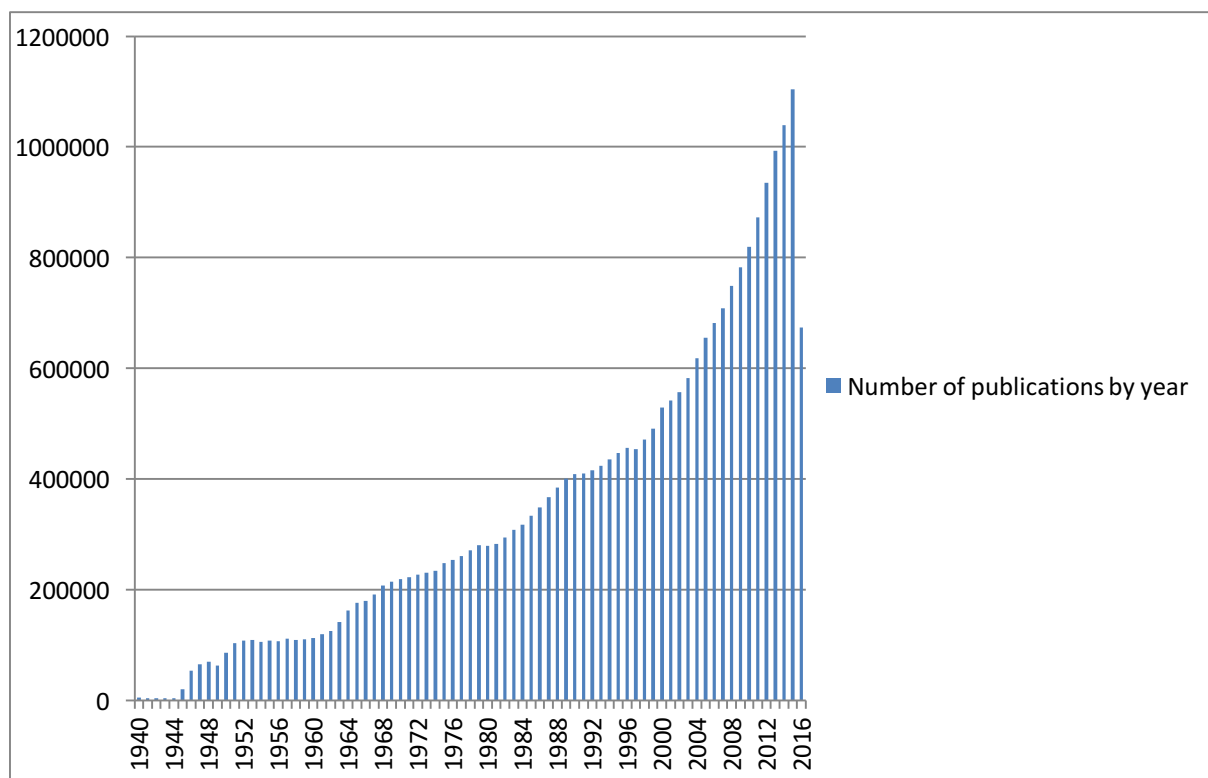


Figure 3: Growth rate of all the articles and the articles related to general practice in PudMed

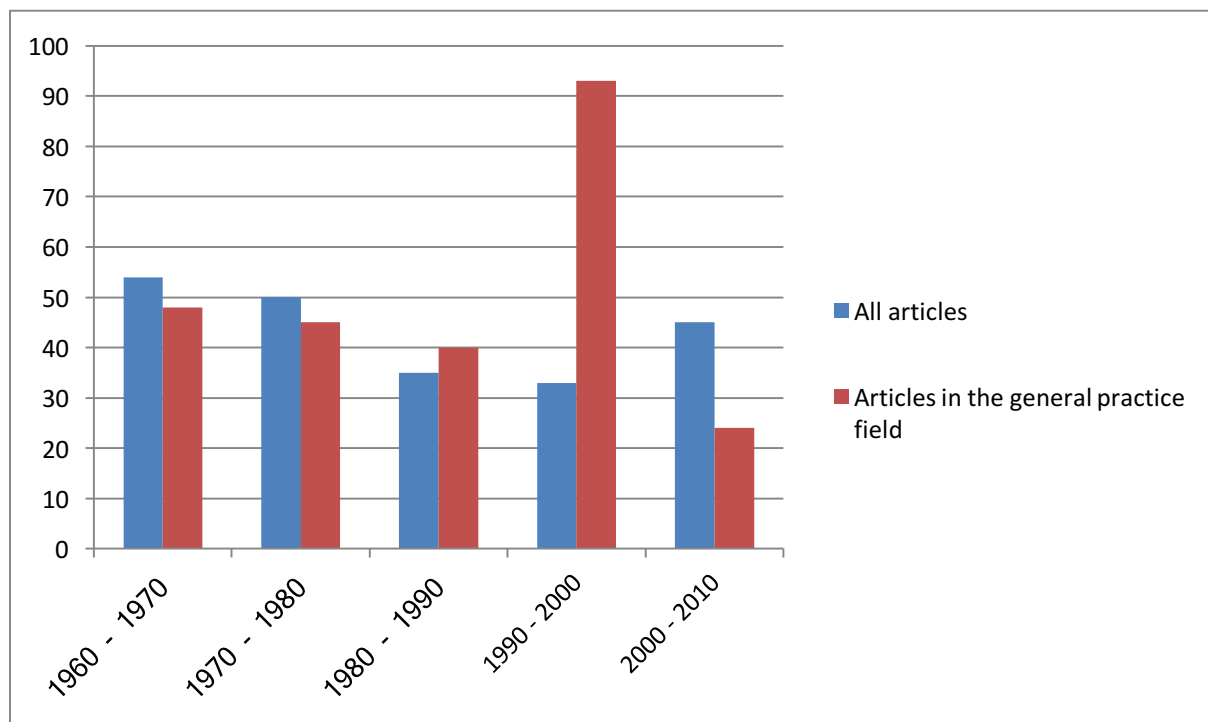


Figure 4: Distribution of articles according to the language

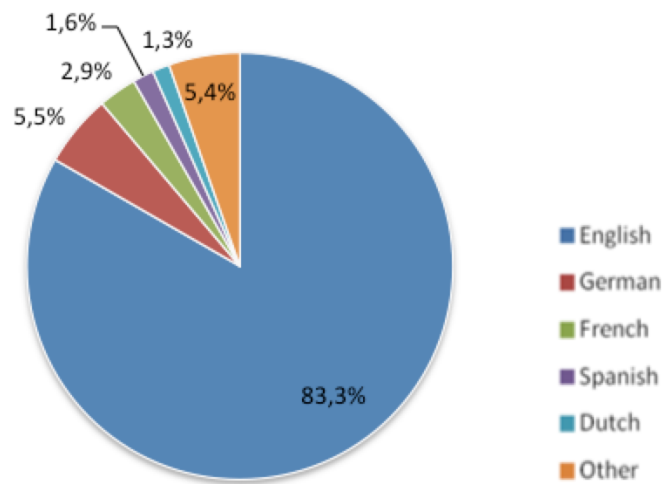
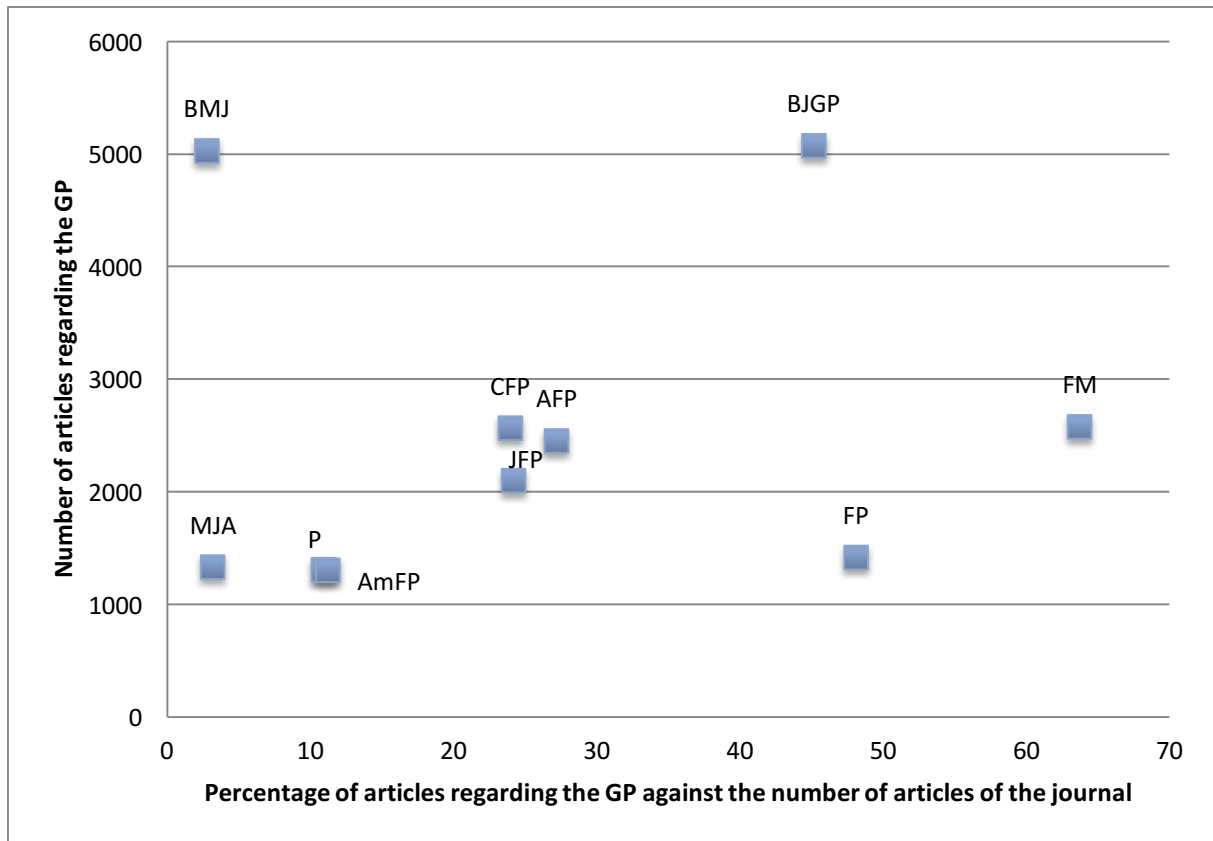


Figure 5: Division of journals according to the number of publications and their coverage field



AFP: Australian Family Physician (Australie)
AmFP: American Family Physician (Etats-Unis)
BJGP: British Journal of General Practice (Royaume-Uni)
BMJ: British Medical Journal (Royaume-Uni)
CFP: Canadian Family Physician (Canada)
FM: Family Medicine (Etats-Unis)
FP: Family Practice (Royaume-Uni)
JFP: The Journal of Family Practice (Etats-Unis)
MJA: Medical Journal of Australia (Australie)
P: The Practitioner (Royaume-Uni)

Table 1: Ranking of countries by number of articles published

Rank	Country	Number of articles (%)
1	USA	9,451 (27.21)
2	UK	7,301 (21.02)
3	Australia	3,269 (9.4)
4	Canada	2,538 (7.3)
5	The Netherlands	2,346 (6.75)
6	Germany	1,349 (3.88)
7	France	879 (2.53)
8	Denmark	815 (2.35)
9	Norway	753 (2.17)
10	Spain	597 (1.72)
11	New Zealand	559 (1.61)
12	Sweden	482 (1.39)
13	Switzerland	469 (1.35)
14	Italy	447 (1.29)
15	Belgium	361 (1.04)
16	Ireland	348 (1)
17	Israel	316 (0.91)
18	India	192 (0.55)
19	Finland	158 (0.45)
20	China	128 (0.37)

Table 2: Ranking of countries by number of articles published related to population size

Rank	Country	Population index per million inhabitants
1	Denmark	143.21
2	Norway	142.83
3	The Netherlands	138.16
4	Australia	134.47
5	New Zealand	122.45
6	UK	112.13
7	Iceland	75.35
8	Ireland	73.82
9	Canada	69.94
10	Switzerland	55.97
11	Sweden	48.92
12	Israel	38.57
13	Slovenia	35.76
14	Belgium	31.74
15	USA	29.16
16	Finland	28.6
17	Malta	28.6
18	Estonia	19.86
19	Croatia	18.93
20	Greenland	17.79

Table 3: Ranking of countries by number of articles published related to GDP per capita

Rank	Country	Number of articles per GDP per capita
1	USA	0.173
2	UK	0.158
3	India	0.121
4	Pakistan	0.058
5	Australia	0.053
6	Canada	0.05
7	The Netherlands	0.045
8	Germany	0.028
9	France	0.021
10	Spain	0.02
11	China	0.017
12	Denmark	0.013
13	Italy	0.013
14	New Zealand	0.013
15	Iran	0.012
16	South Africa	0.012
17	Nepal	0.01
18	Mexico	0.01
19	Nigeria	0.009
20	Israel	0.008

Table 4: Core journals according to Bradford's law

Journal	Impact Factor	Country	Language of publication	Number of GP related articles	Percentage of GP related articles related relative to the total number of published articles
British Journal of General Practice	2.294	UK	English	5,073	45.16
British Medical Journal	17.445	UK	English	5,029	2.78
Family Medicine	1.172	USA	English	2,574	63.74
Canadian Family Physician	1.336	Canada	French/English	2,562	23.97
Australian Family Physician	0.714	Australia	English	2,450	27.21
The Journal of Family Practice	0.894	USA	English	2,096	24.19
Family Practice	1.861	UK	English	1,415	48.1
Medical Journal of Australia	4.089	Australia	English	1,326	3.2
The Practitioner		UK	English	1,308	10.93
American Family Physician	2.175	USA	English	1,300	11.25

Table 5: Ranking of the twenty most cited major MeSH terms

Major MeSH terms	n (%)
Primary health care	4,731 (6.3)
Practice patterns, physicians'	3,553 (4.73)
Attitude of health personnel	3,551 (4.73)
Internship and residency	2,874 (3.83)
Physician-patient relations	2,590 (3.45)
Referral and consultation	2,221 (2.96)
Education, medical, continuing	1,746 (2.32)
Clinical competence	1,698 (2.26)
Physician's role	1,613 (2.15)
Education, medical, graduate	1,493 (1.99)
Education, medical	1,335 (1.78)
Mental disorders	1,157 (1.54)
Health knowledge, attitudes, practice	1,077 (1.43)
Specialization	1,048 (1.39)
State medicine	944 (1.26)
Practice guidelines as topic	920 (1.22)
Hypertension	919 (1.22)
Education, medical, undergraduate	914 (1.22)
Physicians	910 (1.21)
Drug prescriptions	886 (1.18)

6. Conclusion

Les médecins, et en particulier les médecins généralistes, sont confrontés quotidiennement à la difficulté de ne pas tout savoir. En médecine générale, l'étendue du champ d'activité impose aux professionnels une maîtrise de plus en plus aiguë de l'information, qui doit s'appuyer sur le développement d'une littératie numérique, c'est-à-dire la capacité d'utiliser et de comprendre les outils et médias numériques.

Nos travaux ont permis d'identifier plusieurs obstacles à l'accès à la connaissance chez les médecins généralistes. Ces obstacles concernent les professionnels et les outils, certains sont également d'ordre institutionnel. Plusieurs des outils développés par l'équipe du Département d'Informatique et d'Information Médicales du CHU de Rouen ont vocation à répondre à ces difficultés. La base de données bibliographiques LiSSa permet l'accès à plusieurs centaines de milliers de références francophones en santé, issues de PubMed mais également de revues francophones non indexées dans PubMed. Le constructeur de requêtes bibliographiques médicales permet à des utilisateurs non anglophones de construire des équations de recherche complexes dans leur langue maternelle, leur permettant de requêter dans PubMed, LiSSa ou CISMeF.

Malgré leur intérêt, ces outils ne permettront pas de répondre à l'ensemble des besoins des professionnels. A l'instar des principales bases de données bibliographiques existantes, il est probable qu'ils soient considérés par une partie des médecins généralistes comme les éléments d'un modèle académique idéalisé, construit par et pour des chercheurs et des médecins des autres disciplines, éloigné de la réalité de leur champ d'activité. Même pour des professionnels entraînés, il reste difficile d'identifier au sein de ces bases de données les informations pertinentes pour la pratique de la médecine générale. De plus, les professionnels de premier recours plaident pour la mise à disposition d'une information synthétique. Les outils d'aide à la décision existants et accessibles en français sont centrés sur une pathologie ou une population spécifiques. Les outils génériques, comme UpToDate ou Dynamed, sont actuellement disponibles uniquement en langue anglaise.

Les lacunes des professionnels de santé, et notamment des médecins généralistes, dans la maîtrise de la langue anglaise ont constitué l'une des pierres angulaires de ce travail. Avant de réfléchir à la manière d'améliorer la qualité de la formation dans ce domaine, il est important de se poser la question des places relatives de l'anglais et du français dans la transmission de l'information en santé. Les modalités actuelles de valorisation de la recherche et des chercheurs renforcent la place de l'anglais dans le monde de la recherche et appauvrissent progressivement l'espace scientifique francophone. Comme l'ont récemment montré plusieurs revues, comme les Archives de Pédiatrie ou la Revue Neurologique, la langue française est en train de devenir la langue de la formation et non celle de la recherche. Une formation approfondie des professionnels de santé à la langue anglaise ne semble pertinente que s'il apparaît nécessaire à ces mêmes professionnels de santé, non chercheurs, d'avoir accès à des articles de recherche.

Notre travail a également permis de mettre en évidence la nécessité de renforcer le développement d'une culture de la recherche chez les médecins généralistes. Les participants à l'étude qualitative ont à plusieurs reprises évoqué leurs difficultés, voire leur sentiment d'illégitimité, à évaluer la qualité des sources d'information à leur disposition. La diffusion de la culture EBM, l'appropriation par les professionnels des critères de qualité des sources d'information semblent devoir constituer des priorités. L'étendue de leur activité clinique et donc de leur besoin d'information impose aux médecins généralistes une gestion rigoureuse de l'information. Il est probablement nécessaire de favoriser, au sein de la discipline, le développement d'un apprentissage spécifique de la gestion de l'information et des connaissances.

Enfin, le patient reste un acteur incontournable du domaine de l'information en santé. Dans le modèle de l'EBM développé par Sackett, la décision clinique est basée sur les données actuelles de la science, l'expertise clinique du médecin et les besoins et préférences du patient. La construction d'une décision partagée nécessite donc que le patient puisse avoir accès à une information adaptée, compréhensible, indépendante et scientifiquement rigoureuse. Il est donc nécessaire de développer des outils mettant à disposition de telles informations. Par ailleurs, le patient joue également un rôle de plus en plus important dans la production de la connaissance. On peut citer le rôle des objets connectés en santé dans la production de données. Mais en tant qu'expert de la vie avec

la maladie, le patient dispose surtout d'une expérience unique qu'il est nécessaire de prendre en compte. Cela passe notamment par le développement de la participation de patients à l'élaboration et au suivi de projets de recherche en santé.

7. Bibliographie

1. Le concept de compétence appliqué aux compétences clés. Parlement Européen; 2006 Dec.
2. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA. 2002 Jan 9;287(2):226–35.
3. Tardif J. L'évaluation des compétences : Documenter le parcours de développement. Montréal: Chenelière Education; 2006. 363 p.
4. Darmoni SJ, Thirion B, Leroyt JP, Douyère M, Lacoste B, Godard C, et al. A search tool based on “encapsulated” MeSH thesaurus to retrieve quality health resources on the internet. Med Inform Internet Med. 2001 Sep;26(3):165–78.
5. Donnée. In: Wikipédia [Internet]. 2017 [cited 2017 Aug 30]. Available from: <https://fr.wikipedia.org/w/index.php?title=Donn%C3%A9e&oldid=140101826>
6. Document, donnée, information, connaissance, savoir [Internet]. [cited 2017 Aug 30]. Available from: <http://www.les-infostrateges.com/article/031269/document-donnee-information-connaissance-savoir>
7. Association française de normalisation. Vocabulaire de la documentation. Bull Bibl Fr. 1987;(4):386–8.
8. Boubée N, André T. Usages des documents : Qu'est-ce que rechercher de l'information ? : Etat de l'art. Villeurbanne: Ecole Nationale Supérieure Sciences Information Et Bibliothèques; 2010. 286 p.
9. Bibliothèque Interuniversitaire de Santé. La médecine fondée sur les preuves. 2014.
10. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996 Jan 13;312(7023):71–2.
11. Sackett D, Straus S, Richardson W, Rosenberg W, Haynes R. Evidence-Based Medicine: How to Practice and Teach EBM. 2 edition. Edinburgh: Churchill Livingstone; 2000. 280 p.

12. Hodgkin K. Diagnostic vocabulary for primary care. *J Fam Pract.* 1979 Jan;8(1):129–44.
13. Smith R. What clinical information do doctors need? *BMJ.* 1996 Oct 26;313(7064):1062–8.
14. Ely JW, Burch RJ, Vinson DC. The information needs of family physicians: case-specific clinical questions. *J Fam Pract.* 1992 Sep;35(3):265–9.
15. Pham D, Boissel J-P, Wolf P, Rigoli R, Cucherat M, Stagnara J. Médecins généralistes : de quelle information avons-nous besoin ? Une étude quantitative auprès des médecins adhérents à l'Unafarmec RA. *Médecine.* 2008 Oct 1;4(8):369–75.
16. Aubry-Octruc E. L'accès à l'information du médecin généraliste en consultation: enquête auprès des médecins généralistes d'Ile de France : leurs besoins, leurs stratégies de recherche, les sources sollicitées [Thèse d'exercice]. [France]: Université Pierre et Marie Curie (Paris). UFR de médecine Pierre et Marie Curie; 2008.
17. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, et al. Analysis of questions asked by family doctors regarding patient care. *BMJ.* 1999 Aug 7;319(7206):358–61.
18. Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA, et al. A taxonomy of generic clinical questions: classification study. *BMJ.* 2000 Aug 12;321(7258):429–32.
19. González-González AI, Dawes M, Sánchez-Mateos J, Riesgo-Fuertes R, Escortell-Mayor E, Sanz-Cuesta T, et al. Information needs and information-seeking behavior of primary care physicians. *Ann Fam Med.* 2007 Aug;5(4):345–52.
20. Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. *Health Inf Libr J.* 2010 Mar;27(1):2–10.
21. Sources d'information et méthodologie de recherche documentaire. Institut de Recherche et de Documentation en Economie de la Santé; 2016 Apr.
22. La recherche d'informations médicales sur Internet. Haute Autorité de Santé; 2007 May.
23. République Française. Arrêté du 22 décembre 1981.
24. CISMef. CISMef [Internet]. [cited 2017 Aug 30]. Available from: <http://www.chu-rouen.fr/page/base-de-donnees-bibliographiques>

25. Home - PubMed - NCBI [Internet]. [cited 2017 Aug 30]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/>
26. Google Scholar [Internet]. [cited 2017 Aug 30]. Available from: <https://scholar.google.fr/>
27. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. The contribution of databases to the results of systematic reviews: a cross-sectional study. *BMC Med Res Methodol*. 2016 Sep 26;16(1):127.
28. Home | Cochrane Library [Internet]. [cited 2017 Aug 30]. Available from: <http://www.cochranelibrary.com/>
29. Darmoni SJ, Leroy JP, Baudic F, Douyère M, Piot J, Thirion B. CISMeF: a structured health resource guide. *Methods Inf Med*. 2000 Mar;39(1):30–5.
30. Banque de données en santé publique [Internet]. [cited 2017 Aug 30]. Available from: <http://www.bdsp.ehesp.fr/>
31. Haute Autorité de Santé - Page d'accueil - Accueil [Internet]. [cited 2012 Apr 12]. Available from: http://www.has-sante.fr/portail/jcms/j_5/accueil
32. Santé publique France - Accueil [Internet]. [cited 2017 Aug 30]. Available from: <http://www.santepubliquefrance.fr/>
33. FUN - Se former en liberté [Internet]. [cited 2017 Aug 30]. Available from: <https://www.fun-mooc.fr/>
34. www.unf3s.org [Internet]. [cited 2017 Aug 30]. Available from: <http://www.unf3s.org/>
35. Définition et fonctionnalités des portails documentaires | Enssib [Internet]. [cited 2017 Aug 30]. Available from: <http://www.enssib.fr/content/definition-et-fonctionnalites-des-portails-documentaires>
36. Catalogue de bibliothèque. In: Wikipédia [Internet]. 2017 [cited 2017 Aug 30]. Available from: https://fr.wikipedia.org/w/index.php?title=Catalogue_de_biblioth%C3%A8que&oldid=137098800
37. Catalogue SUDOC [Internet]. [cited 2017 Aug 30]. Available from: <http://www.sudoc.abes.fr/>
38. Orphanet [Internet]. [cited 2017 Aug 30]. Available from: <http://www.orpha.net/consor/www/cgi-bin/index.php?lng=FR>

39. The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions [Internet]. [cited 2017 Aug 22]. Available from: <http://www.dlib.org/dlib/december00/weibel/12weibel.html>
40. VIDAL : Base de données médicamenteuse pour les prescripteurs libéraux [Internet]. VIDAL. [cited 2017 Aug 30]. Available from: <https://www.vidal.fr/>
41. BCB Dexther - La base de données sur les Médicaments et les produits de santé [Internet]. [cited 2017 Aug 30]. Available from: <https://www.bcbdexther.fr/>
42. Thériaque [Internet]. [cited 2017 Aug 30]. Available from: <http://www.theriaque.org/apps/contenu/accueil.php>
43. Thésorimed [Internet]. [cited 2017 Aug 30]. Available from: <http://theso.prod-un.thesorimed.org/>
44. OCP&vous [Internet]. [cited 2017 Aug 30]. Available from: <http://www.ocp.fr/>
45. Accueil - Base de données publique des médicaments [Internet]. [cited 2017 Aug 30]. Available from: <http://base-donnees-publique.medicaments.gouv.fr/>
46. Berner E. Clinical decision support systems: State of the Art. Rockville, Maryland.: Agency for Healthcare Research and Quality; 2009 Jun. Report No.: N°09-0060-EF.
47. Séroussi B, Bouaud J. Systèmes informatiques d'aide à la décision en médecine : panorama des approches utilisant les données et les connaissances. Prat Neurol - FMC. 2014 Dec;5(4):303–16.
48. VisualDx - Visual Clinical Decision Support System (CDSS) [Internet]. [cited 2017 Aug 30]. Available from: <https://www.visualdx.com/>
49. Statin Choice Decision Aid - Site [Internet]. [cited 2017 Aug 30]. Available from: <https://statindecisionaid.mayoclinic.org/>
50. UpToDate [Internet]. UpToDate. [cited 2017 Aug 30]. Available from: <http://www.uptodate.com/home>
51. ClinicalKey - Lead with answers [Internet]. [cited 2017 Aug 30]. Available from: <https://www.clinicalkey.fr/#!/>
52. VIDAL - Recommandations [Internet]. [cited 2017 Aug 30]. Available from: <https://www.vidal.fr/recommandations/>
53. Jeanmougin P, Aubert J-P, Le Bel J, Nougairède M. Antibioclic : outil pour une antibiothérapie rationnelle en soins primaires. Rev Prat. 2012 Sep;62:978.
54. GESTACLIC - Accueil [Internet]. [cited 2017 Aug 30]. Available from: <http://gestaclic.fr/>

55. Le suivi des enfants de 0 à 2 ans | Pediadoc [Internet]. [cited 2017 Aug 30]. Available from: <http://pediadoc.fr/>
56. APOROSE | Aide à la Prise en charge de l'OstéopoROSE [Internet]. [cited 2017 Aug 30]. Available from: <http://aporose.fr>
57. Ameli, le site de l'assurance maladie en ligne | assuré | votre compte | ameli.fr [Internet]. [cited 2017 Aug 30]. Available from: <https://www.ameli.fr/assure>
58. Romeyer H. La santé en ligne. Commun Inf Médias Théories Prat [Internet]. 2012 Feb 23 [cited 2017 Aug 30];(Vol. 30/1). Available from: <https://communication.revues.org/2915>
59. Tran Phong E, Allermoz E. Applications en santé : le B.A.-ba sur cette révolution. Médecins Bull Ordre Natl Médecins. 2017 Jul;(50):17–22.
60. Infographie : les usages digitaux des médecins en 2016 [Internet]. Buzz-esanté. 2017 [cited 2017 Aug 30]. Available from: <https://buzz-esante.fr/infographie-usages-digitaux-medecins-2016/>
61. 4e Baromètre VIDAL - CNOM : l'utilisation des smartphones chez les médecins [Internet]. VIDAL France. 2016 [cited 2017 Aug 30]. Available from: <http://www.vidalfrance.com/non-classe/4e-barometre-vidal-cnom-lutilisation-des-smartphones-chez-les-medecins/>
62. Picard R, Vial A. Prospective organisationnelle pour un usage performant des technologies nouvelles en santé. Secrétariat général des ministères chargés des affaires sociales.; 2013 Sep.
63. Les serious games, le futur de la formation médicale ? | esante.gouv.fr, le portail de l'ASIP Santé [Internet]. [cited 2017 Aug 30]. Available from: <http://esante.gouv.fr/le-mag-numero-7/les-serious-games-le-futur-de-la-formation-medicale>
64. IMIA Social Media Working Group [Internet]. IMIA Social Media Working Group. [cited 2017 Aug 30]. Available from: <https://imiasocialmedia.wordpress.com/>
65. Rolls K, Hansen M, Jackson D, Elliott D. How Health Care Professionals Use Social Media to Create Virtual Communities: An Integrative Review. J Med Internet Res. 2016 Jun 16;18(6):e166.
66. Coumou HCH, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. J Med Libr Assoc JMLA. 2006 Jan;94(1):55–60.

67. Ely JW, Osheroff JA, Maviglia SM, Rosenbaum ME. Patient-care questions that physicians are unable to answer. *J Am Med Inform Assoc JAMIA*. 2007 Aug;14(4):407-14.
68. Del Fiol G, Workman TE, Gorman PN. Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA Intern Med*. 2014 May;174(5):710-8.
69. Kosteniuk JG, Morgan DG, D'Arcy CK. Use and perceptions of information among family physicians: sources considered accessible, relevant, and reliable. *J Med Libr Assoc JMLA*. 2013 Jan;101(1):32-7.
70. Bernard E, Arnould M, Saint-Lary O, Duhot D, Hebbrecht G. Internet use for information seeking in clinical practice: a cross-sectional survey among French general practitioners. *Int J Med Inf*. 2012 Jul;81(7):493-9.
71. Cullen RJ. In search of evidence: family practitioners' use of the Internet for clinical information. *J Med Libr Assoc JMLA*. 2002 Oct;90(4):370-9.
72. MacWalter G, McKay J, Bowie P. Utilisation of internet resources for continuing professional development: a cross-sectional survey of general practitioners in Scotland. *BMC Med Educ*. 2016 Jan 21;16:24.
73. Perrin SC. Quels sites internet utilisent les médecins généralistes en consultation ? Enquête de pratique auprès de 100 médecins généralistes libéraux installés en Gironde [Internet]. Université de Bordeaux II; 2016 [cited 2017 Aug 30]. Available from: <https://dumas.ccsd.cnrs.fr/dumas-01278965/document>
74. Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermer JJ, et al. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *BMJ*. 2002 Mar 23;324(7339):710.
75. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. *J Med Internet Res*. 2013 Jun 26;15(6):e122.
76. Vermandel M. IRBM, une nouvelle ère ? *IRBM*. 2012 Sep 25;33(4):241-2.
77. Azulay J-P. Editorial. *Rev Neurol (Paris)*. 2016 Jan;172(1):1.
78. Plaud B, Ashenoune K, Beaulieu P, Beloeil H, Charbit B, Constantin J-M. Ce n'est qu'un au revoir. *Ann Fr Anest Reanim*. 2014;33(12):611-4.
79. Chabrol B, Delacourt C, Sarles J. Archives de Pédiatrie : quelle revue pour demain ? *Arch Pédiatrie*. 2017 Mar;24(3):209-10.

80. Devos P, Dufresne E, Renard JM, Beuscart R. SIGAPS: a prototype of bibliographic tool for medical research evaluation. *Stud Health Technol Inform*. 2003;95:721–6.
81. Bras PL. L'information des médecins généralistes sur le médicament [Internet]. La Documentation Française; [cited 2017 Aug 22]. Available from: <http://www.ladocumentationfrancaise.fr/rapports-publics/074000703/index.shtml>
82. Lebrun L, Fisbach M, Febvre J, Joly-Guillou M-L, Saint-André J-P, Richard I. Le niveau de compréhension de l'anglais des étudiants en médecine peut être amélioré. Résultats d'une stratégie d'évaluation systématique. *Presse Med*. 2011;40(10):471–6.
83. Jamali J, Salehi-Marzijarani M, Ayatollahi SMT. Factors Affecting Journal Quality Indicator in Scopus (SCImago Journal Rank) in Obstetrics and Gynecology Journals: a Longitudinal Study (1999-2013). *Acta Inform Medica*. 2014 Dec;22(6):385–8.
84. Levay P, Ainsworth N, Kettle R, Morgan A. Identifying evidence for public health guidance: a comparison of citation searching with Web of Science and Google Scholar. *Res Synth Methods*. 2016 Mar;7(1):34–45.
85. Henderson J. Google Scholar: A source for clinicians? *Can Med Assoc J*. 2005 Jun 7;172(12):1549–50.
86. Gehanno J-F, Rollin L, Darmoni S. Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Med Inform Decis Mak*. 2013 Jan 9;13:7.
87. Moreira MADM, Lustosa AM, Dutra F, Barros E de O, Batista JBV, Duarte MCS. Public humanization policies: integrative literature review. *Cienc Saude Coletiva*. 2015 Oct;20(10):3231–42.
88. Pascal - Institut de l'information scientifique et technique [Internet]. [cited 2017 Aug 22]. Available from: <http://www.inist.fr/?Pascal>
89. exercer, la revue francophone de Médecine Générale [Internet]. exercer, la revue francophone de Médecine Générale. [cited 2017 Aug 30]. Available from: <https://www.exercer.fr>
90. Thirion B, Loosli G, Douyère M, Darmoni SJ. Metadata element set in a quality-controlled subject gateway: a step to a health semantic Web. *Stud Health Technol Inform*. 2003;95:707–12.

91. Falagas ME, Kouranos VD, Arencibia-Jorge R, Karageorgopoulos DE. Comparison of SCImago journal rank indicator with journal impact factor. *FASEB J Off Publ Fed Am Soc Exp Biol.* 2008 Aug;22(8):2623–8.
92. Sheets L, Gavino A, Callaghan F, Fontelo P. Do language fluency and other socioeconomic factors influence the use of PubMed and MedlinePlus? *Appl Clin Inform.* 2013;4(2):170–84.
93. Fontelo P, Liu F, Leon S, Anne A, Ackerman M. PICO Linguist and BabelMeSH: development and partial evaluation of evidence-based multilanguage search tools for MEDLINE/PubMed. *Stud Health Technol Inform.* 2007;129(Pt 1):817–21.
94. Thirion B, Robu I, Darmoni SJ. Optimization of the PubMed Automatic Term Mapping. *Stud Health Technol Inform.* 2009;150:238–42.
95. Rapport sur la santé dans le monde. Façonner l'avenir. Genève, Suisse: Organisation Mondiale de la Santé; 2003.
96. Starfield B. Primary care: an increasingly important contributor to effectiveness, equity, and efficiency of health services. SESPAS report 2012. *Gac Sanit.* 2012 Mar;26 Suppl 1:20–6.
97. White KL, Williams TF, Greenberg BG. The ecology of medical care. *N Engl J Med.* 1961 Nov 2;265:885–92.
98. Green LA, Fryer GE, Yawn BP, Lanier D, Dovey SM. The ecology of medical care revisited. *N Engl J Med.* 2001 Jun 28;344(26):2021–5.
99. de Pouvoirville G. Développer la recherche en médecine générale et en soins primaires en France: Propositions [Internet]. 2006 [cited 2015 Jul 26]. Available from:
http://lebottinrecherche.fr/sites/default/files/documents/Rapport_De_Pouvoirville.pdf
100. Rosser WW, van Weel C. Research in Family/General Practice is Essential for Improving Health Globally. *Ann Fam Med.* 2004 May;2(Suppl 2):s2–4.
101. Mendis K, Solangaarachchi I. PubMed perspective of family medicine research: where does it stand? *Fam Pract.* 2005 Oct;22(5):570–5.
102. van Weel C. The impact of research in primary care and family medicine: the Thomson Reuters Web of Science Subject Category “Primary Health Care.” *Fam Pract.* 2011 Jun;28(3):239–40.

103. Bertignac C, Fouquet J, Marie E. La Bibliométrie. Université Européenne de Bretagne. 2010. (Form@doct).
104. Baudoin L, Haeffner-Cavaillon N, Pinhas N, Mouchet S, Kordon C. [Bibliometric indicators realities, myth and prospective]. *Med Sci MS*. 2004 Oct;20(10):909–15.
105. Rostaing H. La bibliométrie et ses techniques. Sciences de la société-CRRM. 1996. (Outils et Méthodes).
106. Rosser WW, Starkey C, Shaughnessy R. The pond is wider than you think! Problems encountered when searching family practice literature. *Can Fam Physician*. 2000 Jan;46:103–8.
107. Thirion B, Darmoni SJ. Simplified access to MeSH tree structures on CISMef. *Bull Med Libr Assoc*. 1999 Oct;87(4):480–1.
108. Dunikowski LG, Freeman TR. Impact of family medicine research. *Can Fam Physician*. 2016 Mar;62(3):266–8.
109. Mendis K, Kidd MR, Schattner P, Canalese J. A bibliometric analysis of Australian general practice publications from 1980 to 2007 using PubMed. *Inform Prim Care*. 2010;18(4):223–33.
110. Pshetizky Y, Tandeter H, Tabenkin H, Vinker S, Lahad A, Karkabi K, et al. Thirty years of family medicine publications in Israel (1975-2004): what, where, and how much? *J Am Board Fam Med*. 2009 Feb;22(1):57–61.
111. Merenstein J, Rao G, Wolfe S, Weaver-Agostoni J. Clinical research in family medicine: published articles 2000 and 2005. *Fam Med*. 2009 Jan;41(1):7–8.
112. de Castro FL, Rodríguez OF, Chozas MEM, Hidalgo ER, Lázaro GA. [Research in primary care: 1994-2003]. *Aten Primaria*. 2005 Nov;36(8):415–21.

8. Productions associées à la thèse

1. Publications

- Behaviour and attitudes of residents and general practitioners in searching for health information: from intention to practice. Schuers M, Griffon N, Kerdelhue G, Foubert Q, Mercier A, Darmoni SJ. Int J Med Inform. 2016 May;89:9-14.
- A Search Engine to Access PubMed Monolingual Subsets: Proof of Concept and Evaluation in French. Griffon N, Schuers M, Soualmia LF, Grosjean J, Kerdelhué G, Kergourlay I, Dahamna B, Darmoni SJ. J Med Internet Res 2014;16(11):e271
- Littérature scientifique en santé (LiSSa) : une base de données bibliographiques en français. Griffon N, Schuers M, Kerdelhué G, Grosjean J, Darmoni SJ. Rev Prat. 2017; 67(2) :134-8.
- LiSSa: An alternative in French to browse health scientific literature? Griffon N, Schuers M, Darmoni SJ. Presse Med. 2016 Nov;45(11):955-6.
- LiSSa, Littérature Scientifique en Santé : une base de données bibliographique en français. Griffon N, Rollin L, Schuers M, Douze L, Rodriguez D, Delerue D, Dutoit D, Dahamna B, Kerdelhué G, Grosjean J, Gehanno JF, Darmoni, SJ. Pratique Neurologique - FMC, 2017
- Lost in translation? A multilingual Query Builder improves the quality of PubMed queries. A randomized controlled trial. Schuers M, Joulakian M, Kerdelhue G, Segas L, Grosjean J, Darmoni SJ, Griffon N. BMC Med Inform Decis Mak. 2017 Jul 3;17(1):94

2. Communications orales et affichées

- Accès à l'information médicale : le portail terminologique Francophone CISMef / HeTOP & BDBfr, la base de données bibliographiques en français. Mercier A, Schuers M, Darmoni SJ. Congrès CNGE, Lille, Novembre 2014.
- Interroger Medline en français, c'est possible ! Schuers M, Kerdelhué G, Grosjean J, Darmoni SJ. CMGF, Paris, Mars 2015.
- Outils en ligne en consultation. Gelly J, Jeanmougin P, Schuers M. CMGF, Paris,

Mars 2015.

- Requête en Français dans PubMed : c'est possible et c'est utile ! Griffon N, Schuers M, Kerdelhue G, Grosjean J, Darmoni SJ. 4ème Journée de l'IRIB, Juin 2015
- Une base de données bibliographiques en français pour la Littérature Scientifique en Santé : LiSSa ! Griffon N, Rollin L, Schuers M, Dhombres F, Loiseau S, Douze L, Marcilly R, Chabot O, Delerue D, Dutoit D, Dahamna B, Kerdelhué G, Grosjean J, Gehanno JF, Darmoni SJ. 16e Journées Francophones d'Informatique Médicale. Juin 2016, Genève.
- Thèmes et tendances en médecine générale dans PubMed. Siedlecki C, Griffon N, Kerdelhué G, Grosjean J, Thirion B, Jamoulle M, Darmoni SJ, Schuers M. Congrès CNGE, Grenoble, Novembre 2016.
- Le patient, le médecin et la connaissance. Quels outils au service d'une décision véritablement partagée ? Schuers M, Londres M. CMGF Paris, Avril 2018.