

Université de Montréal

**Construction et validation de filtres de recherche
bibliographique pour soutenir la prise de décisions
basée sur les évidences :
le cas de la Sécurité des Patients**

par

Affaud Anaïs Tanon

Département d'administration de la santé

Faculté de Médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)

en Santé Publique

option organisation des soins

Octobre, 2012

© Affaud Anaïs Tanon, 2012

Université de Montréal
Faculté des études supérieures et postdoctorales

Cette thèse intitulée :

Construction et validation de filtres de recherche bibliographique pour soutenir la prise de décisions basée sur les évidences: le cas de la Sécurité des Patients

Présentée par :
Affaud Anaïs Tanon

a été évaluée par un jury composé des personnes suivantes :

Lambert Farand, président-rapporteur
André-Pierre Contandriopoulos, directeur de recherche
François Champagne, co-directeur
Nicole Leduc, membre du jury
Georges Borgès Da Silva, examinateur externe
Francine Girard, représentant du doyen de la FES

Résumé

Les filtres de recherche bibliographique optimisés visent à faciliter le repérage de l'information dans les bases de données bibliographiques qui sont presque toujours la source la plus abondante d'évidences scientifiques. Ils contribuent à soutenir la prise de décisions basée sur les évidences. La majorité des filtres disponibles dans la littérature sont des filtres méthodologiques. Mais pour donner tout leur potentiel, ils doivent être combinés à des filtres permettant de repérer les études couvrant un sujet particulier. Dans le champ de la sécurité des patients, il a été démontré qu'un repérage déficient de l'information peut avoir des conséquences tragiques. Des filtres de recherche optimisés couvrant le champ pourraient s'avérer très utiles.

La présente étude a pour but de proposer des filtres de recherche bibliographique optimisés pour le champ de la sécurité des patients, d'évaluer leur validité, et de proposer un guide pour l'élaboration de filtres de recherche.

Nous proposons des filtres optimisés permettant de repérer des articles portant sur la sécurité des patients dans les organisations de santé dans les bases de données Medline, Embase et CINAHL. Ces filtres réalisent de très bonnes performances et sont spécialement construits pour les articles dont le contenu est lié de façon explicite au champ de la sécurité des patients par leurs auteurs. La mesure dans laquelle on peut généraliser leur utilisation à d'autres contextes est liée à la définition des frontières du champ de la sécurité des patients.

Mots-clés : Bases de données bibliographiques, Stockage et repérage de l'information, Filtres de recherche bibliographiques, Revue de littérature, Médecine basée sur les faits, Informatique de la santé, Medline, Embase, CINAHL, validation, Sécurité des Patients.

Abstract

Optimized bibliographic search filters are designed to facilitate information retrieval in bibliographic databases, which are almost always the most abundant source of scientific evidence. The purpose of such filters is to support evidence-based decision making. Many of the filters available in the literature are methodological search filters. To reach their full potential they need to be combined with subject filters that identify studies covering a particular topic. In the field of patient safety, it has been demonstrated that deficiencies in the information retrieval process can lead to tragic consequences. Optimized bibliographic search filters covering the field could thus be very useful.

This study is intended to provide subject bibliographic search filters optimized for the field of patient safety, assess their validity, and offer a guide for developing optimized bibliographic search filters.

We propose high-performing bibliographic search filters to retrieve papers dealing with patient safety in health care organizations, that have been explicitly defined as relevant to the patient safety field by their authors, in Medline, Embase and CINAHL. The main generalization issue lies in defining the boundaries of the patient safety field.

Keywords: bibliographic databases, search filters, information storage and retrieval, literature review, evidence-based medicine, public health informatics, Medline, Embase, CINAHL, validity, patient safety.

Table des matières

Chapitre 1 : Introduction	1
La prise de décision basée sur les évidences :	1
Importance de la phase de recherche de l'information :	2
Les obstacles au repérage de l'information dans le domaine de la santé :	3
Repérer les évidences dans les bases de données bibliographiques.....	5
Regard sur les bases de données bibliographiques.	5
Outils pour repérer l'information dans les bases de données bibliographiques.....	7
Les filtres de recherche bibliographiques :	9
Repérer les informations dans le champ de la sécurité des patients :	11
Buts et objectifs de la recherche.....	14
Buts de la recherche:	14
Objectifs spécifiques de la recherche:	15
Chapitre 2 : Méthode de construction, d'évaluation et de validation des filtres de recherche bibliographique optimisés.	16
La construction des filtres de recherche bibliographique optimisés.	16
Description de la méthode :	16
Opérationnalisation de la méthode :	18
Le choix des journaux	18
Constitution d'une liste de journaux dans le domaine de la sécurité des patients	18
Sélection des journaux pour le gold standard:	19
La sélection des articles.	20
La sélection des mots clés.	24
L'évaluation des filtres de recherche bibliographique optimisés.....	25
Sensibilité et Précision :	26
Spécificité et Fallout :	28
Le nombre d'articles à lire et le produit de la sensibilité par la précision	29
La validation des filtres de recherche bibliographique optimisés.....	29
Le choix des bases de données bibliographiques	30

Chapitre 3 : Mise en contexte des résultats et présentation des articles.	32
La construction des filtres de recherche.....	32
Résumé de l'article 1 :	32
Une méthode pour l'évaluation de la validité et de l'utilité des filtres de recherche :.....	33
Résumé de l'article 2 :	33
La validation des filtres de recherche proposés pour Medline, Embase et CINAHL :.....	34
Résumé de l'article 3 :	34
L'utilité des filtres de recherche proposés pour Medline, Embase et CINAHL :.....	35
Résumé de l'article 4 :	36
Un guide pour la construction, l'évaluation et la validation des filtres de recherche bibliographique :	37
Résumé de l'article 5 :	37
Chapitre 4: La construction des filtres de recherche bibliographique.	38
Paper 1: Patient Safety and systematic reviews: finding papers indexed in MEDLINE, EMBASE and CINAHL.....	38
Abstract	38
Introduction.....	39
Method	40
Results.....	43
Discussion	51
Conclusion	52
Competing interests.....	53
Contributors	53
Acknowledgements.....	53
Chapitre 5 : Une méthode pour l'évaluation de la validité et de l'utilité des filtres de recherche bibliographique.....	55
Paper 2: Retrieving information in health bibliographic databases through optimized bibliographic search filters: assessing validity and usefulness	55
Abstract	55
Introduction.....	56

The criterion-related validity framework	61
Assessing the utility of valid search filters	63
Improvement over chance with the chi-squared test of independence	64
Relative improvement over chance using the RIOC statistic.....	68
Criterion-related validity of search filters: preconditions	71
Conclusion	74
Competing interests.....	76
Contributors	76
Acknowledgments.....	76
Chapitre 6: La validation des filtres de recherche proposés pour Medline, Embase et CINAHL.....	77
Paper 3: Patient safety and systematic reviews: A criterion-related validity framework to assess the validity of optimized bibliographic search filters designed for Medline, Embase, and CINAHL	77
Abstract	77
Introduction.....	78
Objective	83
Method	83
The criterion (X)	83
The samples.....	84
The predictors (Y).....	85
The criterion-related validity framework	85
Comparing the two samples.....	87
The correlation	88
Results.....	89
Discussion	92
Conclusion	93
Competing interests.....	94
Contributors	94
Acknowledgments.....	94

Chapitre 7: L'utilité des filtres de recherche proposés pour Medline, Embase et CINAHL.	96
Article 4: Sécurité des Patients et revues systématiques : utilité des filtres de recherche optimisés construits pour Medline, Embase and CINAHL.....	96
Introduction.....	96
Méthode.....	99
Descriptions des gold standards utilisés :	99
Mesurer la validité des sélections faites à l'aide des filtres de recherche.....	102
Tester les filtres de recherche à l'aide du gold standard NPSF.	107
Résultats.....	110
L'analyse de l'utilité des décisions prises à l'aide des filtres de recherche.....	110
Medline.	110
Embase.....	111
CINAHL.....	112
L'analyse de la performance des filtres dans le gold standard NPSF.....	115
La capacité des filtres de recherche à repérer des articles du gold standard NPSF dans la base de données Medline.	115
La capacité des filtres de recherche à repérer des articles du gold standard NPSF dans la base de données Embase.....	117
La capacité des filtres de recherche à repérer des articles du gold standard NPSF dans la base de données CINAHL.	119
Discussion.....	121
Conclusion.....	122
Conflits d'intérêts.....	123
Contributions.....	123
Remerciements.....	124
Chapitre 8 : Un guide pour la construction, l'évaluation et la validation des filtres de recherche bibliographique.....	125
Paper 5: Retrieving scientific papers in bibliographic databases: a step-by-step guide to validated search filters construction.....	125
Abstract.....	125

Introduction.....	126
Method	128
Formulate the question(s).....	129
Define the inclusion and exclusion criteria.....	130
Plan the evaluation of search filters' performance.....	131
Indicators for evaluating search filters.....	131
Collect a set of papers	133
Select the journals	133
Select the databases.....	134
Selecting the gold standard papers.....	135
Select the keywords.....	136
Evaluate and combine the individual search terms.....	137
Validation of the search filters	138
Conclusion	141
Competing interests.....	141
Contributors	141
Acknowledgments.....	141
Chapitre 9: Conclusion.....	143
Retour sur la problématique de départ :	144
Contributions de la thèse :.....	150
Forces et faiblesses :.....	152
Pistes de recherche :	156
Bibliographie.....	157
Annexe 1 : Sélection de journaux pour la recherche manuelle.....	i
Tableau _annexe 1 : Application des trois premiers critères de sélection :	i
La syntaxe utilisée dans Medline pour la sélection définitive des journaux pour la recherche manuelle.....	iii
Annexe 2 : Les formulaires développés pour guider le choix des articles du gold standard.....	iv
Formulaire pour la première étape (Titre et Résumé).....	iv
Formulaire pour la deuxième étape (Textes complets).....	vi

Annexe 3 : Exemple de sélection de termes de recherche.	viii
Les termes commençant par la Lettre S dans les Titres et les Résumés des articles composant le gold standard Composition dans Medline.	viii
Préparation d'un terme sélectionné commençant par S pour interroger la base Medline.	ix
Annexe 4 : The original strategies used by Westwood et al. and their translation into the OVID and EBSCOhost interfaces.	x
Annexe 5 : Search strategies to retrieve patient safety papers in MEDLINE, EMBASE and CINAHL.	xv
Annexe 6 : Number and percentage of papers indexed with the new Mesh "Patient Safety" retrieved by the bibliographic search filters developed for Medline.	xvii
Annexe 7 : Selection guides used to select the gold standard papers.	xviii
Annexe 8 : Caractéristiques des articles du gold standard NPSF en comparaison de ceux du gold standard qui a servi à la construction des filtres de recherche.	xxi
Annexe 9 : Sélection des articles repérés par les filtres de recherche à l'aide d'un tirage aléatoire pour l'estimation de la précision dans le gold standard NPSF.	xxiii

Liste des tableaux

Tableau 1 : Liste initiale de journaux pour la sélection des articles du gold standard.....	19
Tableau 2: Matrice de classification pour l'évaluation des filtres de recherche.	26
Tableau 3 : les indicateurs pour l'évaluation des filtres de recherche.	26
Table 4: Final strategies for Medline (performances of the search strategies in the validation gold standard are presented in parentheses).....	48
Table 5: Final strategies for Embase (performances of the search strategies in the validation gold standard are presented in parentheses).....	49
Table 6: Final strategies for CINAHL (performances of the search strategies in the validation gold standard are presented in parentheses).....	50
Table 7: Classification matrix for evaluating search filters' performance.....	59
Table 8: Indicators to assess the validity of the search filter	63
Table 9: A numerical example of the impact of a validated search filter, using the chi-squared test of independence.	67
Table 10: Formulas to assess the Relative Improvement Over Chance.....	70
Table 11: A numerical example of the impact of a validated search filter using the RIOC statistic	71
Table 12 : Modern views on validity evidence	82
Table 13 : Indicators to assess the validity of the search filter	87
Table 14: Analysis of the samples used to assess criterion-related validity	88
Table 15: Analysis of the predictors: filters developed for Medline, Embase and CINAHL	90
Table 16: Validity coefficients in the two samples used for the validation study	91
Table 17 : Les deux échantillons pour l'analyse de l'utilité.	100
Tableau 18 : Les entités listées dans la publication « Current Awareness » de la US National Patient Safety Foundation entre 1999 et 2008.....	101
Tableau 19 : Formules pour calculer le RIOC (Relative Improvement Over Chance).....	104

Tableau 20 : Statistiques de base, taux de base et taux de sélection pour les filtres de recherche dans l'échantillon 1 qui a servi à la construction des filtres de recherche..	105
Tableau 21 : Statistiques de base, taux de base et taux de sélection pour les filtres de recherche dans l'échantillon 2 qui a servi à l'évaluation de la performance des filtres de recherche.	106
Tableau 22: Matrice de classification pour l'évaluation de la performance des filtres de recherche à l'aide du gold standard NPSF.	108
Tableau 23: Indicateurs pour évaluer la performance des filtres de recherche.....	108
Tableau 24 : Statistiques RIOC pour les filtres de recherche en utilisant les échantillons 1 et 2.....	114
Tableau 25 : Récapitulatif de la performance des filtres de Medline dans les 3 échantillons.	116
Tableau 26 : Récapitulatif de la performance des filtres d'Embase dans les 3 échantillons.	118
Tableau 27 : Récapitulatif de la performance des filtres de CINAHL dans les 3 échantillons.	120
Table 28. Steps for search filter development, assessment and validation	129
Table 29: Classification matrix for the evaluation of search filters	132
Table 30: Indicators for evaluating search filters.....	132
Table 31: Indicators to assess the validity of a search filter	140

Liste des figures

Figure 1 : Schéma de la progression dans SRS.....	23
Figure 2: The conceptual design showing individual terms to be combined for final searches in Medline.....	44
Figure 3: The conceptual design showing individual terms to be combined for final searches in Embase	45
Figure 4: The conceptual design showing individual terms to be combined for final searches in CINAHL.....	46
Figure 5. Percentage of the papers indexed under the MeSH “patient safety” retrieved by the filters developed for Medline	81
Figure 6 : Amélioration de la prise de décision obtenue pour les filtres de Medline.	111
Figure 7 : Amélioration de la prise de décision obtenue pour les filtres d’Embase.....	112
Figure 8 : Amélioration de la prise de décision obtenue pour les filtres de CINAHL.....	113

*Je dédie cette thèse à mes parents : Augou
Brou Christine et Tanon N'Cho Alexis.*

Remerciements

Ma profonde gratitude à toutes les personnes qui m'ont soutenue d'une façon ou d'une autre tout le long de ce projet de doctorat.

Je tiens à remercier tout particulièrement mes deux directeurs de recherche André-Pierre Contandriopoulos et François Champagne qui ont rendu l'aventure du doctorat possible, qui en ont facilité le processus et qui m'ont donné le goût de la recherche. Ils ont multiplié pour moi les opportunités d'initiation au métier de chercheur et étaient toujours là aux moments les plus cruciaux.

Mes sincères remerciements aux professeurs du programme de doctorat en Santé Publique pour la qualité de l'enseignement qui y est dispensé et de façon particulière aux professeurs qui m'ont donné l'opportunité d'être impliquée dans leurs projets de recherche : Jean –Louis Denis, Nicole Leduc et François Béland.

Je remercie les membres de mon jury Lambert Farand, André-Pierre Contandriopoulos, François Champagne, Nicole Leduc, Georges Borgès Da Silva, Francine Girard pour avoir accepté d'évaluer ce travail et pour l'honneur qu'ils me font de participer à mon jury.

Mon parcours de doctorat a été enrichi d'expériences de recherche en dehors des murs de l'université de Montréal qui ont eu un impact important sur ma conception de la recherche et de ses différents débouchés professionnels. Je tiens à remercier toutes les personnes qui ont contribué à cet enrichissement. Plus particulièrement, à l'institut Universitaire Gériatrique de Montréal, ce sont Marie Jeanne Kergoat, Judith Latour, Paule Lebel, Martine Remondin, Bernard Simon-Leclerc et Aline Bolduc, à l'Université Harvard ce sont Michelle Mello, Ashish Jha, Paul Campbell, Pedro Saturno, Michael Reich, Amy Levin et Meri Cayem, à MERLIN, ce sont tout particulièrement Charles Nelson et Arsène Niangoran, au Centre de Recherche de l'Hôpital Charles Lemoyne, Astrid Brousselle et à l'Agence de la Santé et des Services Sociaux de la Montérégie ce sont le Dr Jean Rodrigue, le Dr Jean Ricard, la Dr Sylvie Tardif, et Mme Sylvie Colette.

Je remercie le personnel de l'IRSPUM (Institut de recherche en santé publique de l'Université de Montréal) et tout particulièrement Francine Auger, Josée Tessier, Jean Guy Bouchard, Diane Crevier, Joseph Simigielski. Un grand merci à Linette Saul-Cohen et Nicole Roberge au Département d'Administration de la Santé, et à France Pinsonneault qui jusqu'à sa retraite a été une aide inestimable.

Cette thèse n'aurait pas été possible sans l'appui financier reçu de sources diverses. Je remercie les Instituts canadiens de recherche en santé (IRSC), le programme AnÉIS (Analyse et Évaluation des Interventions en Santé), le GIRU (Groupe de recherche interuniversitaire sur les urgences); le gouvernement de la Côte d'Ivoire, l'Université de Montréal, et le GRIS (Groupe de Recherche interdisciplinaire en Santé).

Je remercie tous les amis qui ont rendu cette aventure plus légère et tout particulièrement Bernadette, Sebastian, Mariam, Annick, Ghislaine et Alphonsine.

Je remercie les membres de ma famille, qui ont été et qui sont toujours là pour moi : mes parents (Christine et Alexis) à qui je dédie cette thèse, mes frères (Christian, Alain, Paul-Henri, Arnaud, Lanzeni, Charles, Yann-Alex, Jeppe) qui sont autant d'anges gardiens, mes nièces (Dominique, Yasmine, Elvire et Maellie) et mon neveu (Pierre-Ilan) qui ensoleillent tous les jours de leur innocence et tout particulièrement ma sœur et meilleure amie Aurélie.

Enfin je remercie Dieu d'avoir mis sur ma route toutes ces personnes, et d'avoir toujours veillé sur moi.

Chapitre 1 : Introduction

La prise de décision basée sur les évidences :

Au cours des dernières décennies, les découvertes scientifiques et les importantes avancées technologiques ont décuplé les possibilités qu'ont les systèmes de santé d'améliorer la santé des populations. Dans le même temps, les systèmes de santé ont fait face à une demande de services de santé qui ne cessait de s'accroître en raison du vieillissement des populations, ainsi que de l'évolution de leurs profils épidémiologiques, de leurs comportements et de leurs attentes.

L'exigence de répondre aux demandes des populations de façon transparente et efficiente, dans un contexte économique de plus en plus difficile, a entraîné un recours plus systématique aux évidences pour orienter les choix et soutenir la prise de décision.

Le mouvement moderne de la «*Médecine basée sur les évidences*» a débuté dans les années 1970 avec les travaux de Archie Cochrane et David Sackett ([Claridge and Fabian 2005](#)). Il a été défini comme "l'intégration des meilleures évidences issues de la recherche avec l'expertise clinique et les valeurs des patients dans la prise de décisions au sujet des soins à délivrer aux patients." ([Sackett, Straus et al. 2000](#)). Très vite, le mouvement a rencontré un succès impressionnant et transformé en profondeur la pratique de la médecine en améliorant les techniques pour améliorer et pour accéder à la base scientifique sur laquelle les décisions cliniques doivent se baser ([Montori and Guyatt 2008](#)). En dépit des nombreuses critiques ([Straus and McAlister 2000](#); [Cohen, Stavri et al. 2004](#)), son énorme influence s'est rapidement répandue de la médecine à d'autres professions de la santé avec la «*Pratique basée sur les évidences* » ([McKibbin 1998](#); [Thomas, Saroyan et al. 2011](#)) et à la santé publique avec la «*Santé Publique basée sur les évidences*» ([Brownson, Gurney et al. 1999](#); [Kohatsu, Robinson et al. 2004](#); [Anderson, Brownson et al. 2005](#); [Muir Gray 2008](#); [Brownson, Chriqui et al. 2009](#); [Brownson, Fielding et al. 2009](#); [Brownson 2011](#)).

Importance de la phase de recherche de l'information :

Au cœur du mouvement des évidences dans le domaine de la santé, se trouve l'exigence pour les chercheurs, les praticiens de la santé et les preneurs de décisions de savoir identifier leurs besoins en information, d'être capable de repérer cette information, de l'évaluer de façon critique et de l'utiliser à bon escient ([Gebbie, Rosenstock et al. 2003](#); [Straus, Richardson et al. 2005](#); [Cobus 2008](#); [Chung and Ram 2009](#); [Thomas, Saroyan et al. 2011](#)). Mais si tous ces éléments sont également importants pour garantir le succès de la démarche, l'étape du repérage des informations est celle sur laquelle nous allons nous appesantir.

Le décès d'une volontaire à un essai clinique menée par l'université John Hopkins en 2001 ([Ramsay 2001](#)) est un tragique rappel de l'importance de cette étape de la prise de décision basée sur les évidences. Lors de l'enquête visant à clarifier les circonstances de ce décès, c'est l'inadéquation de la recherche bibliographique pour construire le protocole de recherche qui a été soulignée ([Steinbrook 2002](#)).

Dans le mouvement de la prise de décision basée sur les évidences, il est important que l'étape du repérage des informations soit la plus complète possible pour assurer la validité des conclusions des études de synthèses et pour éviter les biais ([Egger, Juni et al. 2003](#)). Il est également important que cette étape soit conduite de la façon la plus efficiente possible pour réduire le temps nécessaire à la production et la publication des évidences ([Sampson, Shojania et al. 2008](#)) et de façon générale pour optimiser le temps alloué à cette activité ([Major, Warren et al. 2009](#)).

Malheureusement, une grande proportion des publications ayant pour vocation de synthétiser les évidences ne rapportent pas les stratégies utilisées pour assembler les études primaires ou si elles le font, ne mentionnent pas les preuves de l'efficacité de ces stratégies ([Patrick, Demiris et al. 2004](#); [Golder, Loke et al. 2008](#)).

Les obstacles au repérage de l'information dans le domaine de la santé :

La nature même du champ de la santé publique rend la recherche d'information difficile. Selon Hersh, l'information disponible sous forme digitale de nos jours excède tous les documents produits au cours de 40 000 dernières années de l'histoire humaine ([Hersh 2009](#)). Cela est particulièrement vrai du champ de la santé qui produit une quantité importante d'information ([Harley 2001](#); [Druss and Marcus 2005](#); [Glasziou 2006](#); [Albitz 2007](#); [U.S. National Library of Medicine - National Institutes of Health 2011](#)). En plus d'être abondante, l'information dans le champ de la santé a la caractéristique d'être extrêmement dispersée, ce qui fait que les chercheurs d'information dans le domaine de la santé ont à relever plus de défis que ceux qui évoluent dans un champ plus restreint ([Osioibe 1985](#)). En effet, l'information produite dans le domaine de la santé est disséminée à travers un grand nombre de disciplines, d'organisations professionnelles et de ressources bibliographiques en raison du caractère multidisciplinaire du champ ([Hood and Wilson 2001](#)). Elle est d'autant plus difficile à localiser qu'elle est souvent publiée en dehors des canaux que l'on pourrait qualifier de « naturels ». Une étude bibliométrique a démontré que près de la moitié des articles pertinents en médecine rénale étaient publiés en dehors des journaux de la spécialité ([Garg, Iansavichus et al. 2006](#)), une autre étude a démontré que près de 80% des articles portant sur l'obésité se trouvaient en dehors des journaux de la spécialité ([Baier, Wilczynski et al. 2010](#)). Les chercheurs d'information doivent donc se familiariser avec le langage et la culture de disciplines de la santé autres que leur discipline principale, identifier des chercheurs et des domaines de recherche pertinents à leur recherche et se tenir informés des derniers développements dans les disciplines pertinentes à leur projet ([Hall, Bainbridge et al. 2006](#); [Newby 2011](#)).

Face aux nombreux défis que pose le repérage de l'information dans le champ de la santé, certains ont plaidé que les chercheurs d'information s'allouent les services de bibliothécaires et de professionnels de l'information pour mener à bien leur recherche ([Beverley, Booth et al. 2003](#); [Zipperer 2004](#); [Harris 2005](#); [McGowan and Sampson 2005](#); [Flynn and McGuinness 2011](#)). Il a été démontré que l'assistance des libraires et des professionnels de l'information améliore de façon significative la qualité des recherches

d'information faites par les cliniciens et les chercheurs ([Gardois, Calabrese et al. 2011](#)). Mais disposer de spécialistes de l'information est un luxe que toutes les organisations ne peuvent se permettre ([Kostagiolas, Aggelopoulou et al. 2011](#)). Dans les faits, la plupart des chercheurs, des praticiens et des preneurs de décision conduisent eux-mêmes leurs recherches d'information ([De Groote and Dorsch 2003](#)), ce qui en général produit des recherches d'une faible sensibilité (qui manquent de repérer des ressources pertinentes) ou d'une faible précision (qui repèrent un grand nombre de ressources non pertinentes) ([McKibbon and Walker-Dilks 1995](#); [Grant 2004](#); [Davies 2007](#)).

Par ailleurs, plusieurs études ont démontré que les chercheurs d'information expérimentent de nombreuses barrières non seulement pour accéder à l'information mais aussi pour l'utiliser de façon judicieuse. Ces barrières incluent les compétences des chercheurs d'information, les contraintes de temps, les problèmes que posent les bases de données utilisées, les difficultés à évaluer la pertinence des documents obtenus, etc. ([Hersh, Crabtree et al. 2002](#); [Green and Ruff 2005](#); [Davies 2007](#); [Jackson, Baird et al. 2007](#)). Les chercheurs d'information se sentent souvent débordés par la quantité d'information qu'ils obtiennent à la suite de leur recherche ([Bawden and Robinson 2009](#); [MacDonald, Bath et al. 2011](#)) ou insatisfaits des résultats de leurs recherches ([Davies 2007](#); [Hemminger, Lu et al. 2007](#); [Jackson, Baird et al. 2007](#)).

Repérer les évidences dans les bases de données bibliographiques.

Regard sur les bases de données bibliographiques.

Les bases de données bibliographiques sont l'une des sources les plus abondantes d'information sous forme d'évidences scientifiques ([Haig and Dozier 2003b](#)). Elles peuvent être définies comme des collections numériques organisées de références à la littérature publiée (articles de journaux scientifiques ou de revues, comptes rendus de conférences, rapports, publications gouvernementales et juridiques, brevets, livres ou chapitres de livres, etc.) ([Gwyneth 2003](#)). Elles font partie de la grande famille des systèmes d'information dont le but est de fournir de l'information à une organisation pour l'aider à prendre des décisions ([Swanson 2009](#)). A la différence d'autres systèmes d'information tels que les systèmes d'information et de gestion, les bases de données bibliographiques se caractérisent par le fait que l'information qui y est traitée consiste essentiellement en documents textuels imprimés ou sous forme électronique, plutôt qu'en données chiffrées ([Salton and McGill 1983](#); [Buckland 1997](#)). L'output produit en réponse à une requête faite au système est un ensemble de références qui informent le chercheur d'information sur ce qui a été publié dans un domaine particulier ainsi que sur les sources de ces publications ([Rowley and Hartley 2008](#)). Les bases de données bibliographiques sont, de façon plus spécifique, des formes particulières de systèmes de repérage de l'information, dont le rôle est d'identifier des éléments d'information qu'un utilisateur juge pertinents à l'intérieur de larges collections de documents sous forme de texte, d'images, de vidéos, etc. ([Manning, Raghavan et al. 2008](#); [Larson 2009](#)). Les bases de données bibliographiques se caractérisent par le fait qu'elles sont hautement structurées, ce qui signifie que l'information qu'elles contiennent est très bien définie, de même que le codage utilisé pour représenter cette information ([Meadow 1967](#)), ce qui les différencie d'autres systèmes de repérage d'information sous forme textuelle telles que Google,

Les bases de données bibliographiques sont l'un des premiers champs d'application de la recherche effectuée dans le domaine du repérage de l'information.

La structure des bases de données bibliographiques traditionnelles peut être décrite à l'aide de trois ensembles : un ensemble de documents, un ensemble de requêtes adressées au système par l'utilisateur et un ensemble de mécanismes par lesquels le système fait des liens entre les requêtes qui lui sont faites et les documents qui contiennent l'information recherchée par l'utilisateur. Afin d'optimiser la vitesse à laquelle le système répond à une requête qui lui est faite, les ensembles de termes représentant les documents et ceux représentant les requêtes doivent être classés. La plupart des bases de données bibliographiques d'une certaine taille sont organisées à l'aide d'un index inversé et supportent une logique booléenne pour la construction des filtres de recherche ([Salton and McGill 1983](#); [Manning, Raghavan et al. 2008](#)). Dans les systèmes organisés à l'aide d'un index inversé, le lien entre requêtes et documents n'est pas direct, mais requiert la médiation d'un langage de classification ou d'indexation ([Salton and McGill 1983](#)). Selon les options offertes par le système, ce langage peut contenir des termes spécifiés à l'avance par les concepteurs du système ou des termes libres issus des documents. Des variations existent selon les bases de données bibliographiques quant à la portion du document qui est indexée. Deux processus sont mis en œuvre pour, d'une part, normaliser chacun des documents et les représenter dans le système par un ensemble de termes du langage d'indexation (le processus d'indexation) et pour d'autre part, extraire les documents qui répondent aux requêtes adressées par l'utilisateur (le processus d'extraction). L'index inversé est essentiellement une matrice qui présente tous les documents correspondant à une valeur particulière d'un champ indexé (à l'opposé, l'index direct présente les valeurs de tous les champs indexés pour un document particulier).

Les opérateurs logiques classiques ET, OU et SAUF, de même que des symboles pour signifier la préséance des opérateurs logiques les uns sur les autres sont utilisés dans la construction des filtres de recherche pour repérer les documents. Pour améliorer encore plus la performance de ce genre de systèmes, les fichiers inversés peuvent contenir de l'information sur la position des termes d'indexation les uns par rapport aux autres, et un

poids attribué aux termes selon par exemple un algorithme de calcul de fréquence de termes. Ces options permettent d'utiliser des opérateurs de proximité et de fréquence d'apparition des termes dans la construction des filtres de recherche. Le système peut également contenir des fichiers inversés dédiés au traitement du vocabulaire ce qui permet de prendre en compte les erreurs de typographie et les épellations alternatives dans la construction des filtres de recherche ([Salton and McGill 1983](#); [Korfhage 1997](#); [Manning, Raghavan et al. 2008](#); [Larson 2009](#)).

Outils pour repérer l'information dans les bases de données bibliographiques.

Les chercheurs dans le domaine du repérage de l'information ont développé une longue tradition de l'expérimentation afin de tester des hypothèses sur les meilleures façons de repérer l'information, d'améliorer l'expérience et la satisfaction des utilisateurs et d'évaluer la performance des systèmes mis en place ([Järvelin 2003](#); [Sanderson 2010](#)).

Quelques exemples relativement connus de ces expérimentations sont celles effectuées dans les années 1960 par Cyril Cleverdon et ses collègues (the Cranfield Experiments) pour évaluer l'efficacité de différents systèmes d'indexation, celles menées par Gerald Salton et ses collègues (the SMART (System for the Mechanical Analysis and Retrieval of Text) system) pour étudier de nouvelles approches de repérage de l'information basées sur des modèles où les documents et les filtres de recherche sont représentés dans un espace vectoriel et qui fournissent une liste de documents ordonnés à l'utilisateur, celles initiées par Wilfrid Lancaster dans les années 1965 pour évaluer le système MEDLARS (Medical Literature Analysis and Retrieval System) développé par la NLM (National Library of Medicine), l'ancêtre de Medline (MEDlars onLINE), celles initiées depuis les années 1990 par Dona Harman (TREC : Text Retrieval Conference) pour générer des innovations sur des thèmes aussi divers que les modèles booléens de repérage de l'information, les modèles basés sur un espace vectoriel, les modèles basés sur des réseaux bayésiens, les modèles basés sur le traitement du langage naturel, l'expansion

automatique des filtres de recherche, etc. ([Salton 2009](#); [Tague-Sutcliffe 2009](#); [Voorhees 2009](#); [Sanderson 2010](#)).

Pendant longtemps, un grand nombre des innovations issues de ces différentes expérimentations étaient de simples exercices théoriques. Mais grâce aux avancées dans le domaine de l'informatique puis de l'internet, de nouvelles applications dans le domaine du repérage de l'information ont vu le jour afin de compenser les limitations des systèmes traditionnels basés sur les index inversés à prendre en compte les besoins et les contraintes des utilisateurs ([Croft 1995](#); [Hersh 2009](#)).

Par exemple, dans le domaine biomédical, un outil tel que AliBaba¹ permet de représenter les résultats d'une recherche effectuée dans Pubmed sous forme de graphiques mettant en relation des gènes ou des protéines ([Plake, Schiemann et al. 2006](#)). GoPubmed² est un engin de recherche qui a accès à la base de données Medline, et présente les résultats de recherche sur les gènes ou les protéines sous forme de clusters hiérarchiques ([Doms and Schroeder 2005](#)). Par rapport à Medline, Pubmed est elle-même une interface qui permet l'expansion automatique des filtres de recherche et la présentation des résultats de recherche sous une forme ordonnée³. D'autres applications novatrices dans le domaine biomédical sont décrites dans l'article de [Kim and Rebholz-Schuhmann \(2008\)](#).

Certaines de ces innovations sont plus prometteuses que d'autres. Même si elles sont de plus en plus nombreuses à être disponibles, elles couvrent encore un champ restreint du domaine de la santé. Le recours direct aux bases de données traditionnelles est toujours d'actualité. A côté des initiatives de recherche visant à proposer de nouvelles bases de données, de nouveaux algorithmes de repérage de l'information ou de nouvelles interfaces, un autre groupe de chercheurs propose de développer des filtres de recherche optimisés comme une solution aux difficultés rencontrées par les utilisateurs pour prendre avantage des bases de données bibliographiques organisées selon un index inversé.

¹ <http://alibaba.informatik.hu-berlin.de/about.html>

² <http://gopubmed.org/web/gopubmed/>

³ http://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html

([Haynes, Wilczynski et al. 1994](#); [Greenhalgh 1997](#); [Ford 1999](#); [White, Glanville et al. 2001](#))

Les filtres de recherche bibliographiques :

Les filtres de recherche bibliographique sont les moyens par lesquels un **chercheur d'information** interroge une base de données bibliographiques dans le but de repérer un **ensemble de documents** qu'il juge **pertinent** pour répondre à un **besoin d'information** précis. Ils peuvent encore être définis comme des commandes qui sont la combinaison d'un ou de plusieurs termes de recherche avec d'autres éléments de syntaxe, que l'utilisateur envoie au système pour repérer des documents pertinents parmi ceux indexés dans le système ([Jansen and Rieh 2010](#)).

En réponse à une commande qui lui est faite, le système renvoie une liste non ordonnée d'items qui correspondent de façon exacte à la requête faite par l'utilisateur.

Les filtres de recherche sont donc le résultat de la traduction du besoin d'information exprimé par un chercheur d'information en un langage que reconnaît la base de données bibliographiques. La qualité de cette représentation formelle est limitée par le niveau d'expertise du traducteur, par les options qu'offrent le langage d'indexation et les différents outils de la base de données d'une part et d'autre part par la capacité à appréhender le besoin en information de l'utilisateur ([Naumer and Fisher 2009](#)). En effet, les recherches dans le domaine du repérage de l'information ont démontré que les besoins en information des utilisateurs ne sont pas toujours explicites et qu'ils peuvent même être ignorés du chercheur d'information lui-même ([Taylor 1962](#); [Taylor 1968](#); [Belkin 1980](#); [Belkin, Oddy et al. 1982](#)), qu'ils ne sont pas statiques et peuvent par conséquent évoluer dans le temps et même à l'intérieur d'une session de recherche ([Bates 1989](#)), qu'ils sont enracinés dans le contexte individuel et social de l'utilisateur ([Tuominen and Savolainen 1997](#); [Fisher, Landry et al. 2006](#)). En outre, du point de vue de l'utilisateur, la pertinence d'un document varie en fonction de son niveau de connaissance, du contexte de la

recherche d'information et d'autres facteurs plus ou moins objectifs qu'il est parfois difficile de formaliser ([Saracevic 2007](#); [Saracevic 2008](#); [Cosijn 2009](#); [Howard 2009](#)).

Quand ils sont optimisés, c'est-à-dire construits de façon à maximiser leur performance, les filtres de recherche bibliographique permettent de compenser pour le manque d'habiletés techniques des chercheurs d'information et parfois aussi pour le manque de consistance des méthodes d'indexation dans les bases de données bibliographiques. Ils permettent de limiter une recherche qui par ailleurs aurait produit un volume important de ressources qu'il aurait été difficile de traiter de façon manuelle, d'une façon systématique. Lorsqu'ils sont utilisés à bon escient, ils peuvent réduire le temps et les ressources nécessaires au repérage des informations pertinentes dans les bases de données bibliographiques.

Un grand nombre des filtres bibliographiques optimisés disponibles dans la littérature sont des filtres méthodologiques, construits pour repérer des études conduites à l'aide d'une méthodologie particulière, telle que les essais contrôlés randomisés, les revues systématiques. Le mouvement de la médecine basée sur les faits a largement contribué à populariser ce type de filtres avec les travaux de Haynes et collègues de l'Université McMaster ([White, Glanville et al. 2001](#)). La littérature, propose également de filtres bibliographiques optimisés de contenu, c'est-à-dire des filtres de recherche construits pour repérer la littérature publiée sur un sujet spécifique ([Zlowodzki, Zelle et al. 2006](#); [Iansavichus, Haynes et al. 2010](#)). Ces derniers sont moins courants, et il n'existe pas encore à notre connaissance de filtres de recherche optimisés de contenu pour le champ de la sécurité des patients qui est maintenant une discipline à part entière de la santé publique ([Emanuel, Berwick et al. 2008](#)).

Repérer les informations dans le champ de la sécurité des patients :

Le champ de la sécurité des patients illustre bien les défis qu'il y'a à repérer l'information dans le domaine de la santé. Selon [Stelfox, Palmisani et al. \(2006\)](#), la production scientifique dans le champ de la sécurité des patients a considérablement augmenté depuis la publication du rapport de l'Institute of Medicine (IOM) « To Err is human » ([Kohn, Corrigan et al. 1999](#)). Sur le plan théorique comme sur le plan pratique, de nombreuses disciplines à l'intérieur comme à l'extérieur du champ de la santé, ont été mobilisées pour apporter des solutions pratiques aux problèmes de sécurité des patients ([Miller, Giacomini et al. 2001](#); [Shojania, Duncan et al. 2001](#)). À titre d'exemple, on peut citer la psychologie cognitive, l'ingénierie (facteurs humains), le management organisationnel ([Champagne, Contandriopoulos et al. 2006](#); [Emanuel, Berwick et al. 2008](#)).

Le champ moderne de la sécurité des patients a vu le jour il y'a environ une vingtaine d'années suite aux travaux d'une équipe de chercheurs de l'Université Harvard ([Brennan, Leape et al. 1991](#); [Leape, Brennan et al. 1991](#)) qui avaient pour but de mesurer l'incidence des événements adverses et particulièrement ceux causés par la négligence ou par des soins de mauvaise qualité dans des hôpitaux de New York. A la suite de leurs travaux, ces chercheurs ont établi que les événements adverses se produisent dans environ 3,7% des hospitalisations et qu'environ 27,6% de ces événements adverses étaient causés par de la négligence. Ces résultats relativement inquiétants n'ont pas eu l'écho qu'ils auraient dû avoir à cette époque. C'est la publication du rapport de l'Institute Of Medicine américain « To Err is Human » ([Kohn, Corrigan et al. 1999](#)) qui extrapolait à l'ensemble des États-Unis les résultats des études de l'équipe de Harvard mentionnée plus haut ainsi que ceux d'études similaires effectuées en Utah et au Colorado ([Thomas, Studdert et al. 1999](#); [Thomas, Studdert et al. 2000](#)) qui a entraîné une véritable prise de conscience à travers la planète. Selon ce rapport, entre 44 000 et 98 000 personnes meurent chaque année des suites d'erreurs médicales aux États Unis.

Très vite, de nombreux pays ont pris l'engagement de mobiliser des fonds pour améliorer la sécurité des patients et la qualité des soins. Les résultats d'un tel engagement n'ont pas tardé à se faire sentir. Des institutions ont été créées spécialement pour se pencher sur la question. Des études épidémiologiques ont permis de quantifier le problème à l'échelle nationale dans de nombreux pays ([Vincent, Neale et al. 2001](#); [Davis, Lay-Yee et al. 2002](#); [Davis, Lay-Yee et al. 2003](#); [Baker, Norton et al. 2004](#); [Van den Heede, Sermeus et al. 2006](#); [Michel, Quenon et al. 2007](#)). Au Canada, l'étude réalisée par l'équipe de Peter Norton et Ross Baker a permis d'établir que ce sont entre 9250 et 23750 décès dus aux erreurs médicales qui auraient pu être évitées au cours de l'année 2000 dans les hôpitaux canadiens ([Baker, Norton et al. 2004](#)).

L'intérêt pour la sécurité des patients est passé des hôpitaux à d'autres sites où des soins sont délivrés tels que les soins de santé primaire, les services d'urgence délivrés à l'extérieur des hôpitaux, etc. ([Forster, Murff et al. 2003](#); [Sandars and Esmail 2003](#); [MacDonald, Banks et al. 2008](#); [Thomas and Mackway-Jones 2008](#)).

Depuis la publication du rapport de l'IOM, les initiatives pour élaborer ou mettre en œuvre des interventions visant à améliorer la sécurité des patients se sont multipliées, très souvent sans que la preuve de leur efficacité ne soit faite au préalable. La question de savoir quelle place devait être accordée à la médecine basée sur les évidences dans le jeune champ de la sécurité des patients continue à susciter des débats.

Pour certains auteurs ([Leape, Berwick et al. 2002](#); [Berwick 2005](#); [Berwick 2008](#)), face à l'urgence de la situation, il est important d'agir, même si les connaissances sur lesquelles les interventions sont basées ne répondent pas aux exigences des puristes dans le champ de la médecine basée sur les évidences. Les arguments en faveur de leur position sont entre autres qu'il faut de nombreuses années pour accumuler ce genre d'évidence, que les interventions qui visent le système ne peuvent pas être testées au niveau du patient aussi facilement que les pratiques de soins spécifiques, que la rareté des événements indésirables rend les essais cliniques extrêmement coûteux, que l'accent sur les évidences introduit un biais en faveur des thèmes qui ont suscité l'intérêt des chercheurs et pas nécessairement en faveur des interventions qui semblent importantes sur le terrain. Ils proposent d'élargir le

champ des évidences pour inclure des connaissances issues des méthodes d'apprentissage telles que la « science pragmatique » basées sur l'observation et la pratique réflexive.

Pour d'autres auteurs le paradigme de la médecine basée sur les évidences est né de la prise de conscience que les interventions dans le domaine de la santé, quel que soit le degré auquel elles paraissent plausibles, manquent souvent d'efficacité et parfois même causent des dommages. Puisque les initiatives pour améliorer la sécurité des patients peuvent également se révéler inefficaces, inutiles, voire même nuisibles, il n'y a aucune raison d'exempter la plupart de ces pratiques de l'examen d'une approche fondée sur les preuves ([Shojania, Duncan et al. 2002](#)).

Comme on peut le constater ce qui est considéré comme évidence dans le champ de la sécurité des patients est très large et par conséquent, les sources pour repérer ces évidences sont tout aussi diverses. L'information recherchée peut se retrouver dans plusieurs types de bases de données ou sur le web, néanmoins les bases de données bibliographiques restent une des sources les plus évidentes ([Haig and Dozier 2003a](#)).

Savoir repérer les informations pour soutenir la prise de décision est une compétence clé qui est requise des chercheurs ([Runciman, Baker et al. 2008](#); [Andermann, Ginsburg et al. 2011](#)) et des praticiens ([Ranji and Shojania 2008](#); [Shekelle, Pronovost et al. 2011](#)) dans le domaine de la sécurité des patients. Les filtres de recherche méthodologiques, surtout lorsqu'ils sont optimisés, peuvent contribuer à améliorer la sécurité des patients, notamment en réduisant le temps nécessaire pour produire des revues systématiques d'études répondant à une certaine méthodologie ([Sampson, Shojania et al. 2008](#)). Mais pour donner tout leur potentiel, ils doivent être combinés à des filtres de contenu.

Buts et objectifs de la recherche

Buts de la recherche:

La présente recherche a pour but d'identifier la littérature pertinente pour soutenir la prise de décisions basée sur les évidences dans un champ de la Santé Publique à travers la construction et la validation de filtres de recherche bibliographique.

Le champ de la Sécurité des Patients a été choisi comme un cas particulier, en prélude à une étude plus large ayant pour but d'analyser le concept de sécurité des patients, et plus particulièrement les liens entre Sécurité des Patients et Qualité des Soins qui en dépit des nombreuses avancées effectuées dans le champ de la sécurité des patients au cours des 20 dernières années ne sont pas encore totalement élucidés ([Bates 2008](#); [Emanuel, Berwick et al. 2008](#)). La méthodologie choisie pour cette analyse conceptuelle est une analyse de contenu quantitative de la littérature publiée sur la sécurité des patients. L'analyse de contenu quantitative peut être définie comme une analyse systématique et objective du contenu des messages ([Neuendorf 2002](#)). Elle propose un certain nombre de techniques et d'outils informatiques et statistiques pour examiner de larges collections de textes afin de mettre en évidence des tendances et faire ressortir des connections qui ne seraient pas décelables par les revues systématiques traditionnelles ([He 1999b](#); [Bragge, Relander et al. 2007](#)). Un pré requis à une analyse de contenu quantitative valide est la définition d'une collection de textes qui soit représentative du champ ([Cowie and Lehnert 1996](#); [Stryker, Wray et al. 2006](#); [Hersh 2009](#)). La construction et la validation de filtres de recherche bibliographique contribuent à cet objectif.

Objectifs spécifiques de la recherche:

- Développer des filtres de recherche bibliographique optimisés qui permettent d'identifier un corpus d'articles pertinents dans un champ complexe et diffus tel que le champ de la Sécurité des Patients.
- Évaluer la validité des filtres de recherche bibliographique optimisés développés et sélectionner ceux qui offrent le plus de potentiel de généralisation;
- Proposer un guide pour l'élaboration de filtres de recherche bibliographique optimisés qui puissent être utilisés pour identifier des corpus d'articles pertinents dans d'autres champs de la santé publique.

Chapitre 2 : Méthode de construction, d'évaluation et de validation des filtres de recherche bibliographique optimisés.

La construction des filtres de recherche bibliographique optimisés.

Description de la méthode :

Selon une revue réalisée par [Jenkins \(2004\)](#), il n'y a pas vraiment de consensus sur la meilleure façon de développer des filtres de recherche et quelquefois plusieurs approches sont utilisées de façon concomitante. Néanmoins Jenkins propose une classification en trois groupes basée sur la façon dont les termes de recherche composant les filtres de recherche sont sélectionnés :

- les méthodes de première génération, qui allient les compétences techniques d'un libraire, familier avec les bases de données, aux compétences de contenu des chercheurs qui connaissent leur sujet. Les termes de recherche composant le filtre sont dérivés de façon subjective et on ne propose aucune mesure objective de l'efficacité des filtres développés.
- les méthodes de deuxième génération, qui sont construites comme les premières mais où par contre les filtres développés sont évalués par rapport à un «gold standard», un ensemble de documents pertinents dans le cadre de la recherche, pour déterminer leur efficacité.
- les méthodes de troisième génération, où la recherche d'objectivité commence aux premières étapes de la formation du gold standard. Les termes qui vont composer le filtre de recherche sont par exemple dérivés à partir d'une analyse de la fréquence

de leur apparition dans un ensemble d'articles que l'on juge pertinent. Des méthodes statistiques peuvent être utilisées pour discriminer les termes les plus efficaces de ceux qui le sont moins. On propose non seulement des mesures de l'efficacité des filtres ainsi élaborés, mais encore, une validation de ces filtres à l'aide d'un ensemble indépendant d'articles pertinents. Ces méthodes n'excluent pas une certaine dose de subjectivité, moindre néanmoins que les deux premières.

La méthode adoptée dans le cadre de cette recherche s'inscrit dans la troisième génération. Ce choix se justifie par le souci d'être à la fois le plus exhaustif possible dans la recherche d'articles pertinents au champ de la sécurité des patients, tout en maintenant le volume des documents à traiter dans des proportions raisonnables, afin de proposer aux personnes qui s'intéressent à la Sécurité des Patients un éventail de stratégies efficaces et capables de leur économiser de l'énergie, du temps et des ressources dans leur recherche d'information.

La construction des filtres de recherche s'est donc faite en quatre étapes, décrites plus en détail dans les sections subséquentes :

- Le choix des journaux pour la formation du gold standard;
- La sélection des articles formant le gold standard;
- La sélection des termes composant les filtres de recherche;
- La construction des filtres de recherche à partir des termes les plus pertinents.

Opérationnalisation de la méthode :

Le choix des journaux

Constitution d'une liste de journaux dans le domaine de la sécurité des patients

Dans un premier temps, pour développer les filtres de recherche, une liste de journaux publiant des articles dans le domaine de la sécurité des patients a été construite à partir :

- d'une bibliographie sur le sujet de la sécurité des patients qui nous était familière grâce à des travaux antérieurs sur le thème ([Champagne, Contandriopoulos et al. 2006](#));
- de la bibliographie sur la sécurité des patients publiée par la US National Patient Safety Foundation qui depuis 1999, produit une liste bimensuelle d'articles pertinents sur la sécurité des patients. ([National Patient Safety Foundation](#))
- du répertoire international de périodiques Ulrich (ULRICH'S Periodicals Directory) qui recense plus de 160 000 périodiques et publications en série du monde entier. Ulrich offre la possibilité de faire des recherches de périodique par titre, ISSN, par sujet, par éditeur ou par mots clés, et donne des informations sur l'indexation de ces périodiques dans les bases de données bibliographiques. Les mots clés choisis pour faire la recherche bibliographique dans ULRICH étaient les suivants : medical errors, medication errors, diagnostic errors, adverse events, adverse outcomes, iatrogenic diseases, patient safety.

Nous avons retenu les journaux qui étaient indexés dans MEDLINE et dans au moins une des deux autres bases de données, ce qui a donné une liste de 55 journaux présentés dans Tableau 1.

Tableau 1 : Liste initiale de journaux pour la sélection des articles du gold standard.

Num	Titre	Num	Titre
1.	AACN Clinical Issues.	29.	International Journal of Nursing Practice.
2.	Academic Emergency Medicine.	30.	International Journal of Nursing Studies.
3.	Academic Medicine. 1040-2446.	31.	International Journal for Quality in Health Care.
4.	Accident Analysis & Prevention.	32.	Journal of Clinical Epidemiology.
5.	American Journal of Critical Care.	33.	Journal of Clinical Nursing.
6.	American Journal of Emergency Medicine.	34.	Journal of Critical Care.
7.	American Journal of Epidemiology.	35.	Journal of Evaluation in Clinical Practice.
8.	American Journal of Managed Care.	36.	Journal of Health Economics.
9.	American Journal of Medicine.	37.	Journal of Nursing Administration.
10.	American Journal of Medical Quality.	38.	Journal of Patient Safety.
11.	American Journal of Nursing.	39.	Journal of Quality in Clinical Practice.
12.	American Journal of Public Health.	40.	Journal of Safety Research.
13.	BMC Health Services Research.	41.	Journal of the American Medical Association.
14.	British Journal of Nursing.	42.	Joint Commission Journal on Quality and Safety.
15.	British Journal of Psychology	43.	Lancet, The.
16.	BMJ-British Medical Journal.	44.	Medical Journal of Australia.
17.	Canadian Medical Association Journal.	45.	Medical Letter on Drugs and Therapeutics.
18.	Canadian journal of surgery.	46.	Milbank Quarterly.
19.	Clinical biochemistry.	47.	New England Journal of Medicine.
20.	Critical Care Clinics.	48.	Nursing (year): the voice and vision of nursing
21.	Critical Care Medicine.	49.	Nursing Management.
22.	Effective Clinical Practice.	50.	Nursing Research (IF=1,528)
23.	Family Practice Journal.	51.	Patient Education and Counseling
24.	Frontiers of Health Services Management.	52.	Psychological Medicine
25.	Health Psychology.	53.	Quality Management in Health Care.
26.	Hong Kong Medical Journal.	54.	Quality & Safety in Health Care (IF= 1,937)
27.	International Journal of Clinical Practice.	55.	Quality in health care
28.	International Journal of Health Services.		

Sélection des journaux pour le gold standard:

Pour déterminer la liste finale des journaux à considérer pour la construction du gold standard, 4 critères ont été utilisés :

- Les journaux sont indexés dans l'une au moins des trois bases de données Medline, Embase et CINAHL;
- Les textes complets sont disponibles en ligne;

- Les journaux en sciences infirmières sont représentés;
- Le journal a publié au moins un article sur la sécurité des patients entre 2000 et 2006, selon une vérification rapide effectuée dans la base de données Medline.

L' Annexe 1 présente comment les 55 journaux choisis initialement répondent aux 3 premiers critères ainsi que la syntaxe utilisée dans Medline pour procéder à la sélection finale des journaux. Finalement les journaux sélectionnés pour le gold standard étaient les suivants :

- BMJ (2000);
- JAMA (2001);
- New England Journal of Medicine (2003);
- Quality & Safety in Health Care (2002, 2003, 2004, 2005, 2006);
- American Journal of Nursing (2002);
- Nursing Management (2005, 2006).

La sélection des articles.

Nous avons repéré tous les articles publiés dans les six journaux sélectionnés pendant les années choisies dans les trois bases de données Medline, Embase et CINAHL. Les articles ainsi repérés, au nombre de 11 380 nous ont permis de bâtir une nouvelle base de données.

Les titres et /ou les résumés de ces 11 380 articles ont été importés dans la plateforme SRS de TRIALSTAT ([O'Blenis 2004](#)) pour procéder à la sélection finale des articles qui formeront le gold standard. SRS est une solution web développée par TRIALSTAT pour les revues systématiques. La plateforme permet d'automatiser les différentes étapes d'une revue systématique en plus d'offrir des outils sophistiqués pour l'importation et la gestion des articles.

Vu que l'identification des articles s'est faite simultanément dans les trois bases de données, un grand nombre des articles se retrouvaient en double ou en triple dans la base de données. SRS offre des applications sophistiqués pour identifier et éliminer ces références dupliquées, même si l'indexation dans les bases de données d'origine a entraîné des erreurs dans le nom des auteurs ou le nom des journaux, des erreurs de ponctuation ou d'orthographe, etc...qui pourraient faire penser à des références différentes. En outre, SRS permet de repérer les articles qui ont été publiés à plusieurs reprises à des dates différentes ou dans des journaux différents. Les outils d'identification des références dupliquées de SRS, ont permis de réduire la taille de la base de données de 11380 à 7003 articles.

La sélection des articles composant le gold standard s'est faite en deux étapes selon les critères suivants :

- Les articles doivent mentionner de façon explicite des termes qui renvoient à la sécurité des patients. Le champ de la sécurité des patients étant en pleine évolution, il était important de se concentrer sur les textes dont les auteurs s'inscrivaient de façon délibérée dans le champ de la sécurité des patients par l'utilisation de termes consacrés tels que patient safety, medical errors, medication error(s), diagnostic error(s), adverse event(s), adverse effect(s), adverse outcome(s), undesirable event(s), undesirable effect(s), undesirable outcome(s), medical injury(ies), etc...et les termes qui leur sont proches;
- Les articles parlent explicitement de la sécurité des patients (même s'ils peuvent aborder la sécurité d'autres groupes tels que les prestataires de soins);
- Les articles discutent de la sécurité des patients dans le contexte des organisations de santé.

Les termes sélectionnés pour orienter les réviseurs dans la sélection des articles ont été choisis à l'aide de dictionnaires généraux ou spécialisés, d'encyclopédies et de textes de référence publiés sur la sécurité des patients.

Trois réviseurs (dont deux médecins), ont participé à la sélection et travaillaient de façon totalement indépendante sur la plateforme.

La première étape était très conservatrice : les évaluateurs étaient encouragés à être le moins sévère possible dans l'examen des articles qui était basé uniquement sur le titre et le résumé lorsque disponible. Les trois réviseurs ont participé à cette phase, mais SRS a été configuré de telle sorte que dès qu'un article était examiné par deux réviseurs, il n'était plus disponible pour le troisième réviseur.

- Lorsque deux réviseurs s'accordaient pour estimer que l'article examiné traitait de la Sécurité des Patients dans les organisations de santé, l'article était automatiquement retenu par SRS. Il passait tout de même à la phase 2 pour être examiné une nouvelle fois sur la base du texte complet.
- Lorsque deux réviseurs s'accordaient pour estimer que l'article n'était pas pertinent dans le cadre de la recherche, l'article était automatiquement exclu par SRS et n'était plus disponible pour examen à l'étape 2.
- En cas de conflits (un des réviseurs estimait que l'article devait être retenu et un autre estimait que l'article devait être exclu) ou d'indécision (au moins un des réviseurs n'arrivait pas à se prononcer sur la bases du Résumé et/ ou du Titre de l'article), l'article passait automatiquement à l'étape 2 pour être examiné sur la base du texte complet.

A l'issue de cette première étape, 6535 articles ont été exclus et 468 articles sont passés à la phase 2.

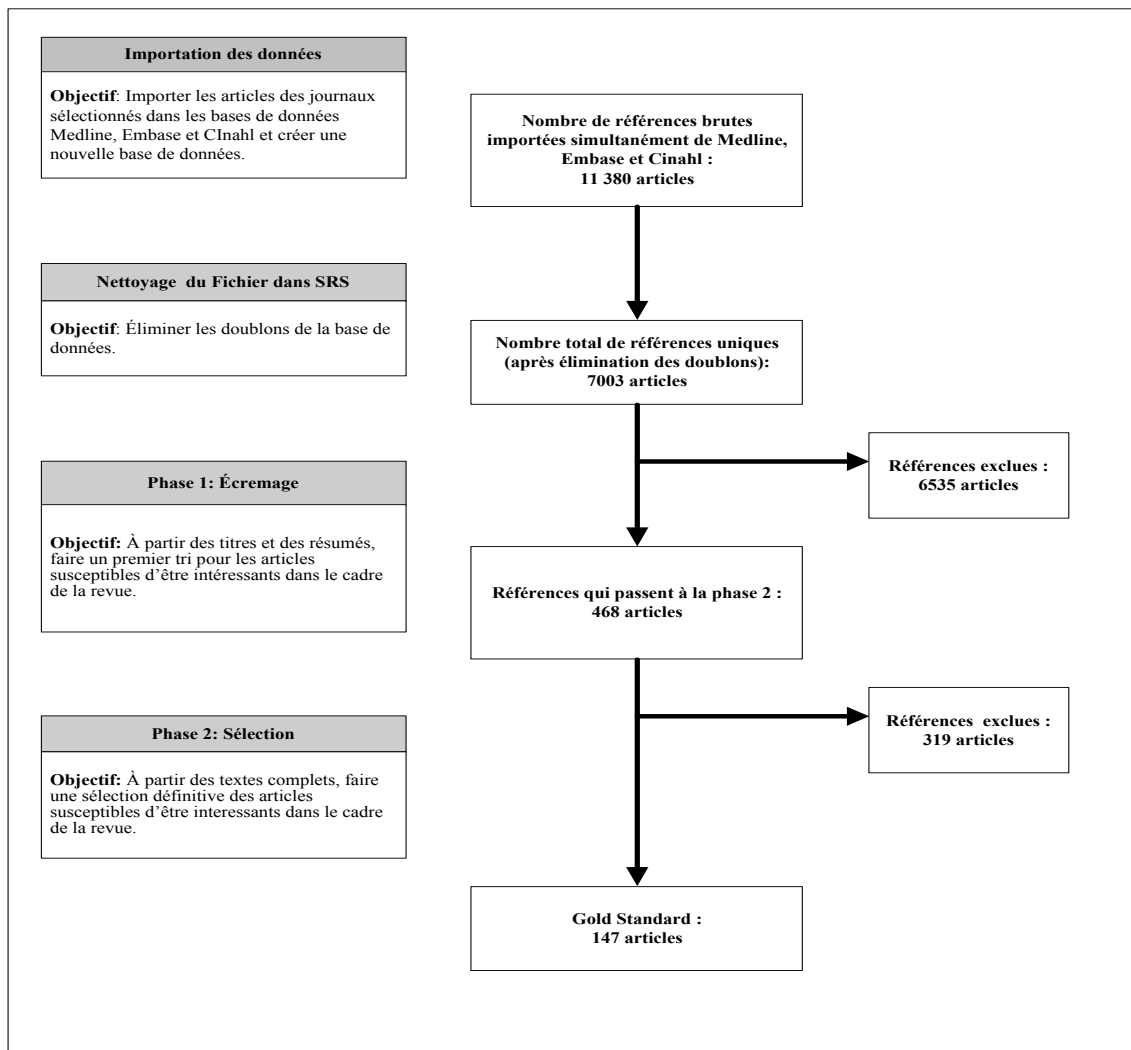
La deuxième étape s'est faite sur la base des textes complets. Les textes complets des 468 articles retenus pour la phase 2 ont été téléchargés dans la plateforme SRS. Après avoir lu le texte complet de l'article, lorsque deux réviseurs répondaient oui à la première question, l'article était retenu. La question 2 forçait les réviseurs à documenter les raisons pour lesquelles ils pensaient que l'article devait être retenu. Lorsque deux réviseurs répondaient non à la première question, l'article était exclu. Tous les articles ont été révisés

de façon indépendante par deux réviseurs. Dans les cas de conflits, l'article était examiné par un troisième réviseur.

L'Annexe 2 présente les formulaires qui ont guidé la sélection des articles à la phase 1 et à la phase 2.

La Figure 1 récapitule le processus de sélection des articles à partir de la plateforme SRS.

Figure 1 : Schéma de la progression dans SRS.



La sélection des mots clés.

Le gold standard a été divisé en deux groupes. L'un des deux groupes a été utilisé pour dériver les mots clés qui serviront à former les filtres de recherche.

Le logiciel Endnote a été utilisé pour générer les listes des mots contenus dans les champs Titre, Résumé et Keyword (Mots clés) pour chacun des groupes d'articles correspondant aux bases de données choisies. Une fois ces listes établies, deux réviseurs ont sélectionné tous les termes potentiellement pertinents, c'est-à-dire qui renvoient à la sécurité des patients dans les organisations de santé. De nouvelles listes ont été élaborées dans Excel à partir de tous les termes pertinents sélectionnés. Nous avons par la suite interrogé les bases de données sélectionnées pour déterminer le nombre d'articles du gold standard dans lesquels ces termes apparaissaient au moins une fois et calculer la valeur des indicateurs (décrits dans la section suivante) retenus pour l'évaluation des filtres de recherche et finalement procéder à la sélection définitive des termes qui serviront à l'élaboration des filtres de recherche. L'Annexe 3 présente un exemple de sélection.

L'évaluation des filtres de recherche bibliographique optimisés

Les indicateurs utilisés pour évaluer la performance des filtres de recherche sont issus d'une longue tradition d'évaluation dans le domaine du repérage de l'information qui remonte aux travaux de Cyril Cleverdon et collègues ([Cleverdon and Keen 1966](#); [Salton and McGill 1983](#)).

Le Tableau 2 présente sous la forme d'une matrice à deux entrées l'échantillon d'articles qui sert à l'évaluation d'un filtre de recherche selon que ces articles font partie ou non du gold standard ou selon qu'ils sont repérés ou non par le filtre évalué. Cette classification permet de définir les ensembles suivants :

- **a** : le nombre de documents pertinents repérés;
- **b** : le nombre de documents non pertinents repérés;
- **c** : le nombre de documents pertinents non repérés;
- **d** : le nombre de documents non pertinents non repérés;
- **(a+b)** : l'ensemble des documents repérés;
- **(c+d)** : l'ensemble des documents non repérés;
- **(a+c)** : l'ensemble des documents pertinents;
- **(b+d)** : l'ensemble des documents non pertinents;
- **N** : le nombre total de documents dans l'échantillon.

Tableau 2: Matrice de classification pour l'évaluation des filtres de recherche.

		Échantillon d'articles		Total
		Pertinent (gold standard)	Non Pertinent	
Filtre de recherche	Articles repérés	a	b	(a+b)
	Articles non repérés	c	d	(c+d)
Total		(a+c)	(b+d)	N

Le Tableau 3 présente les formules des cinq indicateurs que nous avons retenus pour l'évaluation des filtres de recherche.

Tableau 3 : les indicateurs pour l'évaluation des filtres de recherche.

Indicateurs	Formule	Explication
Sensibilité	$\frac{a}{(a+c)}$	$\frac{\text{Nombre d'articles pertinents repérés par le filtre de recherche}}{\text{Nombre total d'articles composant le Gold Standard}}$
Précision	$\frac{a}{(a+b)}$	$\frac{\text{Nombre d'articles pertinents repérés par le filtre de recherche}}{\text{Nombre total d'articles repérés par le filtre de recherche}}$
Spécificité	$\frac{d}{(b+d)}$	$\frac{\text{Nombre d'articles non pertinents repérés par le filtre de recherche}}{\text{Nombre total d'articles non pertinents}}$
Sensibilité* Précision	$\frac{a^2}{(a+c)(a+b)}$	Indicateur mesurant l'équilibre entre la sensibilité et la précision
Nombre d'articles à lire	$\frac{(a+b)}{a}$	Nombre d'articles non pertinents à lire en moyenne avant de trouver un article pertinent

Sensibilité et Précision :

La sensibilité et la précision sont les deux indicateurs les plus souvent utilisés pour comparer les filtres de recherche ([Hersh 2009](#); [Sanderson 2010](#)). Elles mettent toutes les deux l'attention sur les vrais positifs. La sensibilité est la proportion des documents pertinents qui sont repérés par les filtres de recherche ($\text{Sensibilité} = \frac{a}{a+c}$), tandis que la précision est la proportion des documents repérés par les filtres de recherche qui est pertinente ($\text{Précision} = \frac{a}{a+b}$).

Dans les bases de données bibliographiques traditionnelles, les filtres de recherches sont des combinaisons booléennes de mots clés. Chaque filtre de recherche produit un ensemble non ordonné (en termes de pertinence) de documents qui sont présentés à l'utilisateur. Ces documents s'avèrent être soit pertinents pour l'utilisateur, soit non pertinents. Par conséquent pour chacun de ces filtres de recherche bibliographique et si l'on considère un ensemble fermé de documents, il est possible de calculer une seule précision et une seule sensibilité ([Salton and McGill 1983](#)). D'autres systèmes de repérage de l'information permettent, en fonction de l'algorithme qui établit des correspondances entre les requêtes soumises par les utilisateurs et la collection de documents, de présenter à ces derniers des quantités variables de documents repérés, ordonnés en fonction de la pertinence. Il est alors possible d'obtenir différentes paires de valeurs pour la sensibilité et la précision, en fonction de la quantité d'information qui est présentée à l'utilisateur ([Salton and McGill 1983](#)). À partir de ces paires sensibilité- précision, il est possible de construire une courbe sensibilité-précision. Les courbes sensibilité-précision sont orientées vers l'utilisateur des systèmes d'information qui sont en règle générale particulièrement intéressés à maximiser le repérage des documents pertinents. ([Salton and McGill 1983](#)).

Dans l'un ou l'autre de ces types de bases de données bibliographiques, les études empiriques ont démontré qu'il existe un arbitrage entre la sensibilité et la précision : plus on améliore la sensibilité et plus la précision a tendance à diminuer et vice versa ([Buckland and Gey 1994](#)). La relation inverse entre la sensibilité et la précision des filtres de recherches était à l'origine une hypothèse de travail découlant du projet Cranfield I au cours duquel Cyril Cleverdon et son équipe ont testé l'efficacité de systèmes d'indexation ([Aslib. Cranfield Research Project. and Cleverdon 1962](#)). Elle était énoncée comme suit : « *There is an inevitable inverse relationship between recall and relevance* » ([Cleverdon \(1964\)](#), p. 2).

Suite aux critiques ([Swanson 1965](#)) et aux résultats d'autres expérimentations ([Salton 1965](#)) cette affirmation a été reformulée comme suit : « *Within a single system, assuming that a sequence of sub searches for a particular question is made in the logical order of expected decreasing precision, and the requirements are those stated in the*

question, there is an inverse relationship between recall and precision, if the results of a number of different searches are averaged » ([Cleverdon 1972](#)).

Autrement dit, rien ne s'oppose en théorie comme en pratique à ce que, pour une instance de recherche d'information, l'on soit capable d'élaborer un filtre de recherche qui tout en améliorant soit la sensibilité soit la précision, soit également capable d'améliorer l'autre mesure ([Salton 1965](#); [Salton and McGill 1983](#)), néanmoins dans un système donné et lorsque l'on considère un grand nombre de filtres de recherche, la précision a tendance à diminuer à mesure que la sensibilité augmente et vice versa ([Rowley and Hartley 2008](#)).

Spécificité et Fallout :

La spécificité ($Spécificité = \frac{d}{b+d}$) mesure la proportion des documents non pertinents qui ne sont pas repérés par les filtres de recherche. En échangeant les cas positifs pour les cas négatifs dans le Tableau 2 et si on cherche à prédire les cas négatifs, on peut dire que la spécificité est le pendant inverse de la Sensibilité ([Powers 2011](#)).

Le Fallout est le complément à 1 de la Spécificité ($Fallout = \frac{b}{b+d} = 1 - Spécificité$). Dans le domaine du repérage de l'information, ce dernier indicateur permet de mesurer la performance du système pour les documents non pertinents de la même façon que la sensibilité mesure la performance du système pour les documents pertinents ([Salton and McGill 1983](#)).

Lorsque ces indicateurs sont calculés pour un système qui renvoie une liste de documents ordonnés selon leur pertinence, il est possible d'obtenir différentes paires de valeurs pour la sensibilité et le Fallout ou encore $(1 - Spécificité)$, en fonction de la quantité d'information qui est présentée à l'utilisateur. Le graphique construit à partir de ces différentes paires (Sensibilité; Fallout) est connu sous le nom de la courbe ROC (*Receiver Operating Characteristics*). Les courbes sensibilité-fallout sont orientées vers le système d'information car elles indiquent avec quelle efficacité les documents non pertinents sont rejetés. ([Salton and McGill 1983](#)).

Nous avons retenu la Spécificité comme indicateur, car c'est celui des deux indicateurs que l'on retrouve le plus souvent dans les articles portant sur les filtres de recherche bibliographiques dans le domaine de la santé

Le nombre d'articles à lire et le produit de la sensibilité par la précision

Le nombre d'articles à lire est l'inverse de la précision. Cet indicateur donne de l'information sur le nombre d'articles qu'il est nécessaire de lire avant de trouver un article pertinent repéré par le filtre de recherche. ([Bachmann, Coray et al. 2002](#)). Le produit de la sensibilité par la précision est un autre indicateur qui permet de faire un arbitrage entre sensibilité et précision ([Bachmann, Coray et al. 2002](#)).

La validation des filtres de recherche bibliographique optimisés

L'importance de la validation des filtres de recherche a été soulignée à plusieurs reprises, mais n'est pas encore une pratique courante ([Jenkins 2004](#); [Glanville, Bayliss et al. 2008](#)). Le constat général est que les filtres de recherche bibliographique, conçus dans un environnement expérimental ont tendance à être moins performants lorsque appliqués en situation réelle ([Jenkins 2004](#)). Pour cette raison, il est recommandé aux concepteurs de filtres de recherche bibliographique de procéder à la « validation externe » des filtres qu'ils proposent de façon à donner aux utilisateurs potentiels, une idée de ce à quoi ils peuvent s'attendre en termes de performance dans des conditions réelles d'utilisation. La méthode généralement préconisée pour cette « évaluation externe », consiste à mesurer la performance des filtres de recherche dans un gold standard différent de celui qui a servi lors de la construction des filtres de recherche ([Glanville, Bayliss et al. 2008](#)).

Pour répondre à cette exigence, nous avons construit un nouveau gold standard à partir des suggestions de lecture proposées par La National Patient Safety Foundation une organisation américaine créée en 1997 pour améliorer la sécurité des patients. Deux fois par mois depuis 1997, la NPSF-US publie une liste annotée de ressources pertinentes sur la

sécurité des patients, nommée « Current Awareness » ([National Patient Safety Foundation](#)) ([National Patient Safety Foundation resource page](#)).

Mais mesurer la performance des filtres de recherche dans un gold standard externe ne permet pas de tirer des conclusions pertinentes sur la validité des filtres de recherche et le terme « validation externe » convient plus aux devis de recherche. La validation externe réfère à la mesure dans laquelle les relations entre deux concepts mises en évidence dans le cadre d'une étude peuvent être généralisées à d'autres contextes de recherche ([McDonald 2005](#)). Parce que les filtres de recherche sont des instruments de mesure, nous proposons dans le cadre de cette thèse que la validité soit plutôt évaluée à l'aide du devis de la validation de critère. Nous avons choisi de présenter le modèle élaboré dans la section résultat de la thèse, plutôt que dans la section Méthode pour illustrer d'une part le fait que c'est à l'issue de plusieurs essais erreurs que la nécessité de repenser la validation des filtres de recherche bibliographique s'est imposée à nous et d'autre part, le fait que cette méthode reste encore à peaufiner.

Le choix des bases de données bibliographiques

Nous avons choisi de travailler avec les bases de données Medline, Embase et CINAHL car ces trois bases de données sont parmi les plus populaires utilisées dans le domaine de la santé. ([Haig and Dozier 2003a](#); [Fineout-Overholt, Hofstetter et al. 2005](#)).

Medline (**Medical Literature Analysis and Retrieval System Online**) est la base de données bibliographiques en sciences biomédicales dont la collection de références est la plus grande au monde ([Huber, Boorkman et al. 2008](#)). Elle est produite par la Bibliothèque nationale de médecine des États-Unis d'Amérique (National Library of Medicine). Elle contient plus de 19 millions d'articles provenant de plus de 5000 journaux. La couverture commence à partir de 1946 (avec quelques ressources plus vieilles). Le matériel indexé provient majoritairement de journaux académiques. Près de 91% du matériel indexé est écrit en anglais. Une caractéristique qui distingue Medline est que le matériel est indexé à

l'aide d'un vocabulaire contrôlé (Medical Subject Headings ou Mesh). ([U.S. National Library of Medicine - National Institutes of Health 2012](#)).

CINAHL (**The Cumulative Index to Nursing and Allied Health Literature**) complète Medline en couvrant les disciplines des sciences infirmières et des sciences paramédicales. Par rapport à Medline, elle répertorie en plus des articles des journaux académiques, des ressources telles que les chapitres de livres, des livres, des actes de conférences, des thèses, des documents audiovisuels, etc. ([Pickett 2008](#)). Tout comme Medline, CINAHL a un langage d'indexation ([Lawrence 2007](#)).

Embase produite par Elsevier complète Medline avec une plus grande couverture des sources européennes, et également des ressources dans le domaine de la pharmacie et des médicaments. Embase a plus de 5 millions de titres qui ne sont pas offerts par Medline. ([Woods and Trewheellar 1998](#); [Elsevier 2012](#))

Chapitre 3 : Mise en contexte des résultats et présentation des articles.

La construction des filtres de recherche

Le **chapitre 4** largement inspiré de la revue réalisée par [Jenkins \(2004\)](#), présente sous forme d'article notre opérationnalisation de la méthode de construction de filtres de recherche de 3^{ème} génération ainsi que l'évaluation des filtres de recherche optimisés que nous avons développés pour repérer les articles portant sur la sécurité des patients dans les bases de données Medline, Embase et CINAHL ([Tanon, Champagne et al. 2010](#)).

Résumé de l'article 1 :

Objectif: élaborer des filtres de recherche pour identifier les documents sur la sécurité des patients dans Medline, Embase et CINAHL.

Méthode: Nous avons parcouru la version électronique de six journaux afin d'identifier des articles portant sur la sécurité des patients, répondant à nos critères de sélection, et publiés entre 2000 et 2006. Les articles sélectionnés ont été divisés en deux gold standards, l'un pour construire les filtres de recherche et le deuxième pour les valider. Les termes de recherche potentiels pour la construction des filtres de recherche ont été identifiés à l'aide d'une analyse de fréquence des mots contenus dans les titres, les résumés et les mots clés utilisés pour indexer les documents dans les bases de données. Les recherches ont été effectuées à l'aide de chacun des termes sélectionnés dans chacune des bases de données afin d'identifier les articles indexés dans les bases de données sous ces termes. La sensibilité, la précision et la spécificité ont été calculées pour chacun des termes sélectionnés. Les termes avec une sensibilité supérieure à 10% ont été combinés pour former les stratégies finales. Les stratégies de recherche développées ont été exécutées dans l'échantillon d'article correspondant au gold standard de validation afin d'évaluer leur performance. Une dernière étape dans le processus de validation était de comparer la performance de chaque stratégie à celle d'autres stratégies trouvées dans la littérature.

Résultats: nous avons développé des stratégies pour les trois bases de données qui sont très sensibles (de 95% à 100%), précises (entre 40% et 60%) et équilibrées (le produit de la sensibilité et la précision étant de l'ordre de 30% à 40%). Les stratégies sont très spécifiques et ont surclassé celles trouvées dans la littérature portant sur le même sujet.

Conclusion: les stratégies que nous avons développées peuvent répondre aux besoins des utilisateurs visant à maximiser soit la sensibilité ou la précision, ou la recherche d'un compromis raisonnable entre la sensibilité et la précision, lors de la recherche d'articles portant sur la sécurité des patients dans Medline, Embase ou CINAHL.

Une méthode pour l'évaluation de la validité et de l'utilité des filtres de recherche :

Le **chapitre 5** présente sous forme d'article la méthode que nous avons élaborée pour mesurer la validité des filtres de recherche. La littérature sur les filtres de recherche bibliographique aborde la validation des filtres en termes de validation externe, un terme qui n'est pas bien adapté à la nature des filtres de recherche qui sont des instruments de mesure. Nous soutenons dans ce chapitre que l'évaluation de la validité de critère convient mieux aux filtres de recherche et proposons un modèle qui s'inspire de la tradition psychométrique.

Résumé de l'article 2 :

Les filtres de recherche bibliographique sont de plus en plus populaires dans le domaine de la santé comme des outils permettant d'extraire efficacement des documents dans des bases de données bibliographiques. La méthode utilisée pour établir leur validité, qui consiste à évaluer la performance des filtres dans un gold standard externe, donne souvent des résultats contradictoires qui font penser qu'en réalité, les filtres de recherche ne sont pas très utiles. Toutefois, la validité ne consiste pas à reproduire en conditions non expérimentales, les mêmes performances obtenues lors de la construction des filtres de

recherche. Elle consiste plutôt à s'assurer que l'instrument mesure ce qu'il est censé mesurer et que les inférences faites sur un groupe spécifique de sujets suite à l'utilisation de l'instrument sont appropriés pour un usage donné. Le cadre d'évaluation de la validité de critère qui reflète le cadre utilisé pour développer les filtres de recherche bibliographique et évaluer leur performance offre une alternative plus appropriée pour évaluer la validité de l'interprétation des filtres de recherche et leur utilité. Ce cadre, inspiré par le champ de la psychométrie, devrait intégrer de façon cohérente les contraintes imposées par les différents types de filtres de recherche bibliographique, leurs caractéristiques spécifiques en tant qu'instruments de mesure, ainsi que la structure et les caractéristiques offertes par les bases de données bibliographiques pour lesquelles les filtres de recherche ont été construits afin d'être plus efficace.

La validation des filtres de recherche proposés pour Medline, Embase et CINAHL :

Le **chapitre 6** présente sous forme d'article la validité de critère des 27 filtres de recherche bibliographique développés pour les bases de données Medline, Embase et CINAHL et évaluée selon le modèle proposé dans le chapitre 5.

Résumé de l'article 3 :

Objectif: évaluer la validité de critères des filtres de recherche bibliographique élaborés pour Medline, Embase et CINAHL afin de repérer des articles qui portent sur la sécurité des patients dans les organisations de santé.

Méthode: Les données ont été recueillies à l'aide des échantillons utilisés pour élaborer et évaluer la performance des filtres de recherche dans les trois bases de données. La validité concomitante a été évaluée en calculant les coefficients Phi.

Résultats: Les coefficients Phi ont été rapportés à leur valeur maximale afin de compenser la restriction de leurs étendues et faciliter l'interprétation. La corrélation avec le

critère était très élevée dans les deux échantillons et dans toutes les bases de données, pour les stratégies sensibles ainsi que pour les stratégies équilibrées de Medline ($\phi / \phi_{\max} \geq 0,90$). Dans les deux échantillons, elle était élevée pour la plupart des stratégies équilibrées de Embase et de CINAHL ($0,70 \leq \phi / \phi_{\max} < 0,89$) et modérée pour les stratégies précises de Embase ($0,50 \leq \phi / \phi_{\max} < 0,69$). Elle variait de faible à modérée pour la plupart des stratégies précises dans les bases de données Medline et CINAHL en fonction de l'échantillon considéré ($0,30 \leq \phi / \phi_{\max} < 0,69$).

Conclusion: À notre connaissance, cette étude est la première du genre à présenter la validité de critère des filtres de recherche bibliographique. Les filtres de recherche les plus sensibles et équilibrés peuvent valablement être interprétés comme une mesure de ce que l'article repéré porte sur la sécurité des patients dans les organisations de santé. Sur la base de ces résultats, les auteurs encouragent davantage de recherche sur les méthodes d'échantillonnage appropriées pour la construction et la validation des filtres de recherche bibliographique ainsi que l'utilisation d'instruments fiables et exhibant une bonne validité de contenu pour la sélection du gold standard. Les stratégies les plus sensibles devraient être considérées en premier lorsqu'on a l'intention d'appliquer les filtres de recherche en dehors du contexte expérimental de leur construction.

L'utilité des filtres de recherche proposés pour Medline, Embase et CINAHL :

Le **chapitre 7** présente sous forme d'article une mesure de l'utilité des 27 filtres de recherche élaborés pour les bases de données Medline, Embase et CINAHL par le biais de l'évaluation de la validité des sélections que permettent ces filtres de recherche, ainsi qu'une mesure de la performance des filtres dans un ensemble de données assemblé selon une définition plus large de la sécurité des patients.

Résumé de l'article 4 :

Objectif: (1) Évaluer l'utilité de 27 filtres de recherche qui ont été élaborés pour repérer des articles portant sur la sécurité des patients dans les organisations de santé, ancrés de façon explicite dans le champ de la sécurité des patients par leurs auteurs, dans les bases de données Medline, Embase et CINAHL. (2) Évaluer dans quelle mesure ces filtres pourraient être utilisés pour repérer des articles où le thème de la sécurité des patients est abordé selon une perspective plus large.

Méthode: (1) La statistique du RIOC (Relative Improvement Over Chance), qui mesure l'efficacité de la prédiction des instruments de mesure tout en corrigeant pour la chance et les écarts importants entre les valeurs moyennes des variables arrangées dans un tableau de contingence à double entrée, a été utilisée pour mesurer le pourcentage de bonnes décisions prises en utilisant chacun des filtres de recherche qui excède le pourcentage des bonnes décisions qui auraient pu être prises en sélectionnant les articles de façon aléatoire. Deux échantillons d'articles ont été utilisés. (2) Les filtres de recherche ont été testés à l'aide d'un gold standard externe comprenant des articles compilés selon une définition de la sécurité des patients plus large que celle qui a prévalu lors de la construction des filtres de recherche.

Résultats: Les stratégies les plus sensibles et les plus équilibrées dans chacune des trois bases de données et dans les deux échantillons améliorent la sélection des articles répondant à nos critères de plus de 85% par rapport à une sélection aléatoire (la valeur du RIOC est supérieure à 85%). Les stratégies les plus précises améliorent la qualité de la prise de décision au-delà de la chance dans une proportion qui varie entre 41% et 83%. Certaines stratégies se distinguent parce qu'elles semblent plus utiles dans un échantillon que dans un autre. Lorsqu'appliqués à un gold standard qui reflète une définition plus large de la sécurité des patients, les filtres les plus sensibles et les plus équilibrés ont une sensibilité qui varie entre 85% et 42%, entre 57% et 37% et entre 60% et 37% respectivement pour Medline, Embase et CINAHL.

Conclusion: Les stratégies que nous avons développées sont très utiles pour repérer les articles portant sur la sécurité des patients dans les organisations de santé dans les bases

Medline, Embase et CINAHL. Ces filtres ne couvrent pas tout le champ de la sécurité des patients tel que le démontre leur sensibilité dans le nouveau gold standard. Néanmoins en l'absence de filtres plus adaptés à des définitions plus larges de la sécurité des patients, les filtres proposés peuvent être utiles, même dans ce contexte, pour débiter une revue systématique.

Un guide pour la construction, l'évaluation et la validation des filtres de recherche bibliographique :

Le **chapitre 8** reprend toute la démarche que nous avons adoptée pour construire les filtres de recherche bibliographique en intégrant les leçons que nous avons apprises le long du processus.

Résumé de l'article 5 :

Les professionnels de la santé, les décideurs et les chercheurs doivent acquérir les compétences nécessaires pour identifier leurs besoins d'information et localiser, évaluer et utiliser plus efficacement cette information afin de prendre avantage des possibilités offertes par les derniers développements technologiques dans les secteurs de l'information et des communications. L'une de ces compétences est la capacité à développer des filtres de recherche pour récupérer des documents qui répondent à des besoins d'information spécifiques dans les bases de données bibliographiques. Malgré le fait qu'on retrouve de plus en plus dans la littérature des descriptions de la méthode adoptée pour développer des filtres optimisés de recherche bibliographique, il n'existe à notre connaissance, aucun guide compact non seulement pour le développement mais aussi pour la validation des filtres de recherche bibliographique. Par conséquent, nous proposons un guide basé sur notre expérience, et optimisé grâce à une revue de la littérature afin d'aider les chercheurs en santé, les décideurs et les praticiens dans leurs activités de recherche d'information.

Chapitre 4: La construction des filtres de recherche bibliographique.

Paper 1: Patient Safety and systematic reviews: finding papers indexed in MEDLINE, EMBASE and CINAHL.

Affaud Anaïs Tanon, François Champagne, André-Pierre Contandriopoulos, Marie-Pascale Pomey, Alain Vadeboncoeur, Hung Nguyen.

Abstract

Objective: to develop search strategies for identifying papers on patient safety in Medline, Embase and CINAHL.

Methods: six journals were electronically searched for papers on patient safety published between 2000 and 2006. Identified papers were divided into two gold standards, one to build and the other to validate the search strategies. Candidate terms for strategy construction were identified using a word frequency analysis of titles, abstracts and keywords used to index the papers in the databases. Searches were run for each one of selected terms independently in every database. Sensitivity, precision and specificity were calculated for each candidate term. Terms with sensitivity above 10% were combined to form the final strategies. The search strategies developed were run against the validation gold standard to assess their performance. A final step in the validation process was to compare the performance of each strategy to those of other strategies found in the literature.

Results: we developed strategies for all three databases that were highly sensitive (range 95% to 100%), precise (range 40% to 60%) and balanced (the product of sensitivity and precision being in the range of 30% to 40%). The strategies were very specific and outperformed those found in the literature.

Conclusion: the strategies we developed can meet the needs of users aiming to maximize either sensitivity or precision, or seeking a reasonable compromise between sensitivity and precision, when searching for papers on patient safety in Medline, Embase or CINAHL.

Introduction

After the Institute of Medicine in the U.S. released its 1999 report *To Err Is Human*, publications on patient safety proliferated.([Stelfox, Palmisani et al. 2006](#)) Given the growing emphasis on evidence-based practice and decision-making, accessing these resources is an important factor in improving patient safety.([Zipperer 2004](#)) However, end users—researchers, health professionals, decision-makers—encounter several barriers to retrieving and using this knowledge: difficulties in accessing journals, lack of time for an efficient search, inadequate technical skills, database complexity, cost of information searches, etc.([McKibbon and Walker-Dilks 1995](#); [Green and Ruff 2005](#); [Davies 2007](#))

This is why many authors strongly recommend engaging a librarian for information searches.([King 1987](#); [Marshall 1992](#); [Zipperer 2004](#); [Weightman and Williamson 2005](#)) However, with the increased availability of electronic databases, most users conduct their own bibliographic searches ([De Groote and Dorsch 2003](#); [Grant 2004](#)), generally producing results with poor sensitivity (missing relevant items) or precision (with a great number of irrelevant items). ([McKibbon and Walker-Dilks 1995](#); [Grant 2004](#); [Davies 2007](#))

To optimize electronic database searches, many authors advise using search filters. ([Haynes, Wilczynski et al. 1994](#); [Greenhalgh 1997](#); [White, Glanville et al. 2001](#))

These combine text words (or terms in the title and abstract) and specific index terms to retrieve relevant articles or documents of a given type or on a given subject in a particular database. Thus, the literature increasingly offers search filters developed to facilitate the process of retrieving information that could otherwise take months to develop and would mobilize significant resources.([Golder, McIntosh et al. 2006](#)) Most are methodological filters whose aim is to retrieve studies carried out using a specific

methodology, such as randomised controlled trials.[\(Grant 2004\)](#) The methods used to develop them vary greatly. First-generation filters are derived subjectively with no objective measure of their efficacy. Filters derived from second-generation methods are constructed the same way, but evaluated against a “gold standard”—a set of relevant records used to test efficacy. In third-generation methods, the quest for objectivity begins in the first stages of developing the gold standard, continues with measurements of filter efficacy, and culminates in a validation using an independent set of relevant records.[\(Jenkins 2004\)](#)

This article offers, to those interested in patient safety in the context of healthcare processes, effective search filters developed using third-generation methodology that can save time, effort and resources when searching for papers on patient safety, i.e., academic publications explicitly mentioning the term “patient safety” or dealing with subjects such as medical, diagnostic or medication errors, adverse or undesirable events, effects or outcomes, or medical injuries, in three databases: Medline (OVID interface), Embase (OVID interface) and CINAHL (EBSCOhost interface).

Method

The search filters were constructed and validated in several stages.

First, to develop the gold standard, journals were selected from a corpus of articles on patient safety known to the authors from their previous work in the field. This initial list was expanded to include journals from the extensive bibliography on the National Patient Safety Foundation’s website,ⁱ as well as from Ulrich’s Periodicals Directory,ⁱⁱ which lists international journals on many subjects, including patient safety.

Of the 55 journals thus listed, we retained six (BMJ (2000), JAMA (2001), New England Journal of Medicine (2003), Quality & Safety in Health Care (2002-2006), American Journal of Nursing (2002), and Nursing Management (2005-2006)) using these criteria: the journal is indexed in either Medline, Embase, or CINAHL; complete texts were easily available online, the journal published at least one article on patient safety between

2000 and 2006, as verified by a quick search in Medline,ⁱⁱⁱ and nursing journals were represented.

In a second phase, the journals selected were simultaneously searched in Medline, Embase and CINAHL. The search produced 11 380 articles, which were imported into TrialStat's SRS platform for systematic reviews,^{(O'Blenis 2004)^{iv}} an IT solution which automates the workflow and data management aspects of systematic reviews through the screening and data abstraction phases. Two research assistants separately reviewed the 7003 records left after eliminating duplicates. An initial screening of these articles based on titles and abstracts eliminated 6535 articles that did not explicitly deal with patient safety. After a full reading of the remaining 468 articles, 147 that dealt specifically with patient safety in the healthcare system were retained and constituted our gold standard. Because they were not uniformly listed in all three databases, the size of the gold standard varied: Medline (147), Embase (140), and CINAHL (108).

To allow us to validate the search filters with journals not used in their construction, the gold standard was divided into two convenience samples: a "composition gold standard" using JAMA (2001), Quality & Safety in Health Care (2004–2005 and 2006), and Nursing Management (2005), and a "validation gold standard" using BMJ (2000), New England Journal of Medicine (2003), Quality & Safety in Health Care (2002–2003), and Nursing Management (2006). After in-depth screening, no relevant reference was found in the American Journal of Nursing (2002). The composition gold standard was made up of 83, 79 and 56 references from Medline, Embase and CINAHL, respectively, while the validation gold standard had 64, 61 and 52.

In a third phase, EndNote^v files containing the articles of the composition gold standard for each database were created to identify terms to retain in constructing the search filters: free text from article titles or abstracts and specific index terms used by the databases to classify the documents. Lists of terms from each source (titles, abstracts, keywords) were created using Endnote's "define term lists" tool. Two reviewers examined these lists independently to select all potentially relevant terms and then classified them according to a conceptual design that would guide their AND/OR linkages. Any

disagreements between the two assessments were resolved by a third reviewer. Three concepts constituted the skeleton of this design: “PATIENT” for terms referring to patients, “SAFETY” for terms referring to safety and “CARE” for terms referring to care processes.

New lists were developed in Microsoft Excel to ascertain the number of gold standard articles in which these terms appeared at least once, to calculate a certain number of indicators and to do the definitive selection of the terms that would be used to develop the search filters.

Five indicators were used to evaluate the individual terms and final searches:

- Sensitivity: $(\text{Number of } \mathbf{relevant} \text{ articles retrieved by the search}) / (\text{TOTAL number of relevant articles in the universe of articles (gold standard)}) * 100.$
- Precision: $(\text{Number of } \mathbf{relevant} \text{ articles retrieved by the search}) / (\text{TOTAL number of articles } \underline{\text{extracted}} \text{ by the search}) * 100.$
- Specificity: $(\text{Number of } \mathbf{non-relevant} \text{ articles } \underline{\text{not retrieved}} \text{ by the search}) / (\text{Number of non-relevant articles in the universe of articles}) * 100.$
- Sensitivity*Precision: the product of sensitivity and precision, which allows us to reach equilibrium between sensitivity and precision.
- Number of articles to read: $1/\mathbf{PRECISION}$, which allows us to identify the number of non-relevant articles that must be read in order to find one relevant article.

To validate the search filters, we used three distinct procedures. First, we measured the performance of filters developed in the validation gold standard to evaluate their internal validity, i.e., the integrity of the approach used to develop them. Second, an important step in our validation process consisted of verifying whether there were already high-performance search strategies in the literature for retrieving articles on patient safety. Westwood et al.’s systematic review for the National Health Service (NHS) published in 2002 presented various strategies used in several databases ([Westwood, Rodgers et al. 2002](#)) to retrieve studies on patient safety. Their definition of patient safety was close to ours and, like us, in the search strategies they proposed for Medline, Embase and CINAHL

they put no restrictions on the type of publication retrieved. We reproduced these strategies as faithfully as possible in the OVID and EBSCOhost interfaces (see Annexe 4) to retrieve papers in our two gold standards. Finally, to establish external validity, a new gold standard was generated from readings suggested by the National Patient Safety Foundation between 1999 and 2008. This is a corpus of articles entirely different from those of the study and covering several more years. This gold standard will allow us to determine the extent to which our filters can be generalized. The results of this last phase will be presented in a second article.

Results

All individual terms with >10% sensitivity were retained for inclusion in combinations to be used in constructing final search strategies. Figures 1, 2 and 3 present the individual terms that were chosen based on the conceptual design for Medline, Embase and CINAHL. Figures for sensitivity, precision and specificity are indicated, respectively, in parentheses after each term.

By combining terms using “OR”, our goal was to achieve search strategies with 100% sensitivity and 100% specificity, and by using “AND”, to achieve at least 50% precision.

Figure 2: The conceptual design showing individual terms to be combined for final searches in Medline

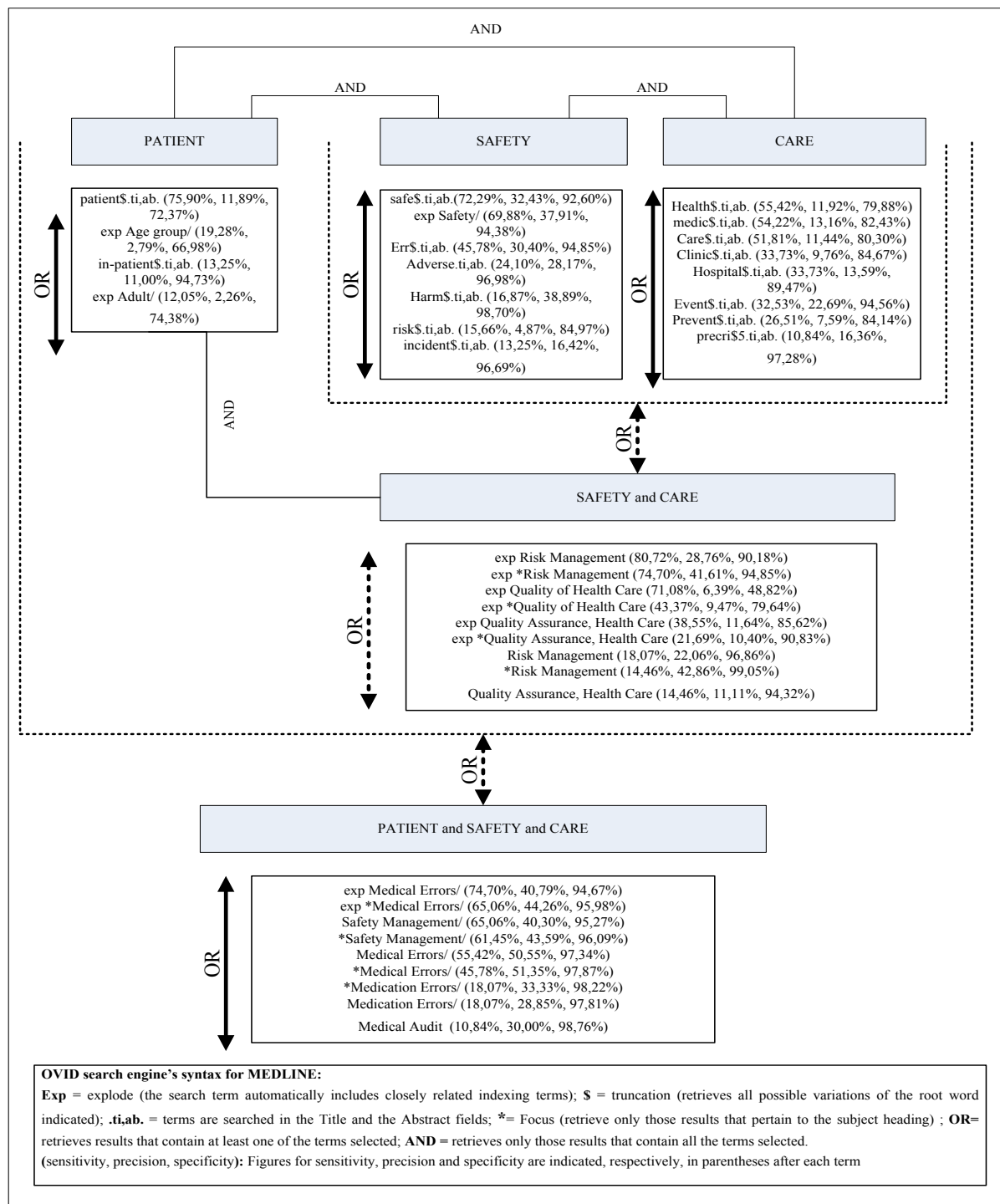


Figure 3: The conceptual design showing individual terms to be combined for final searches in Embase

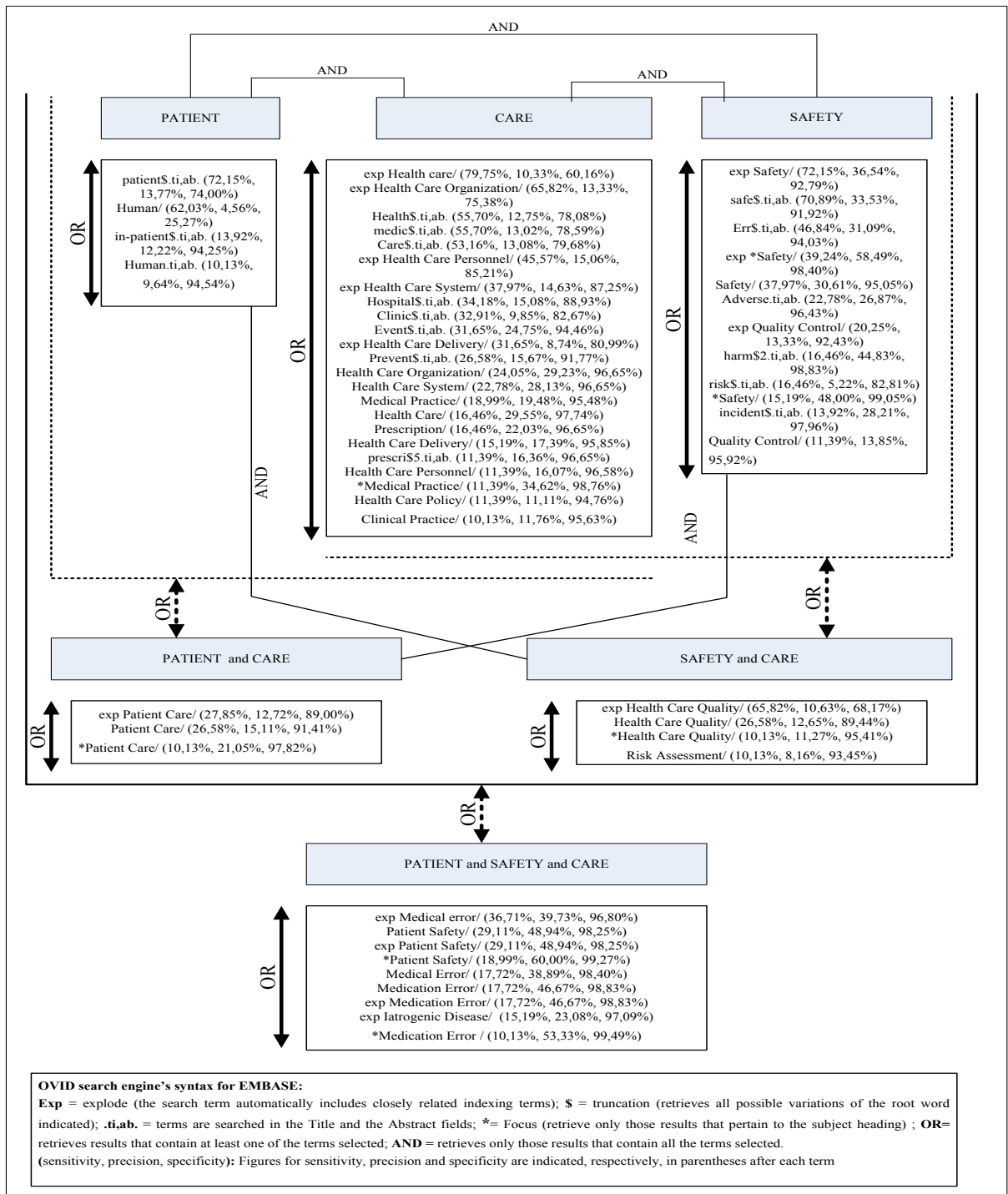
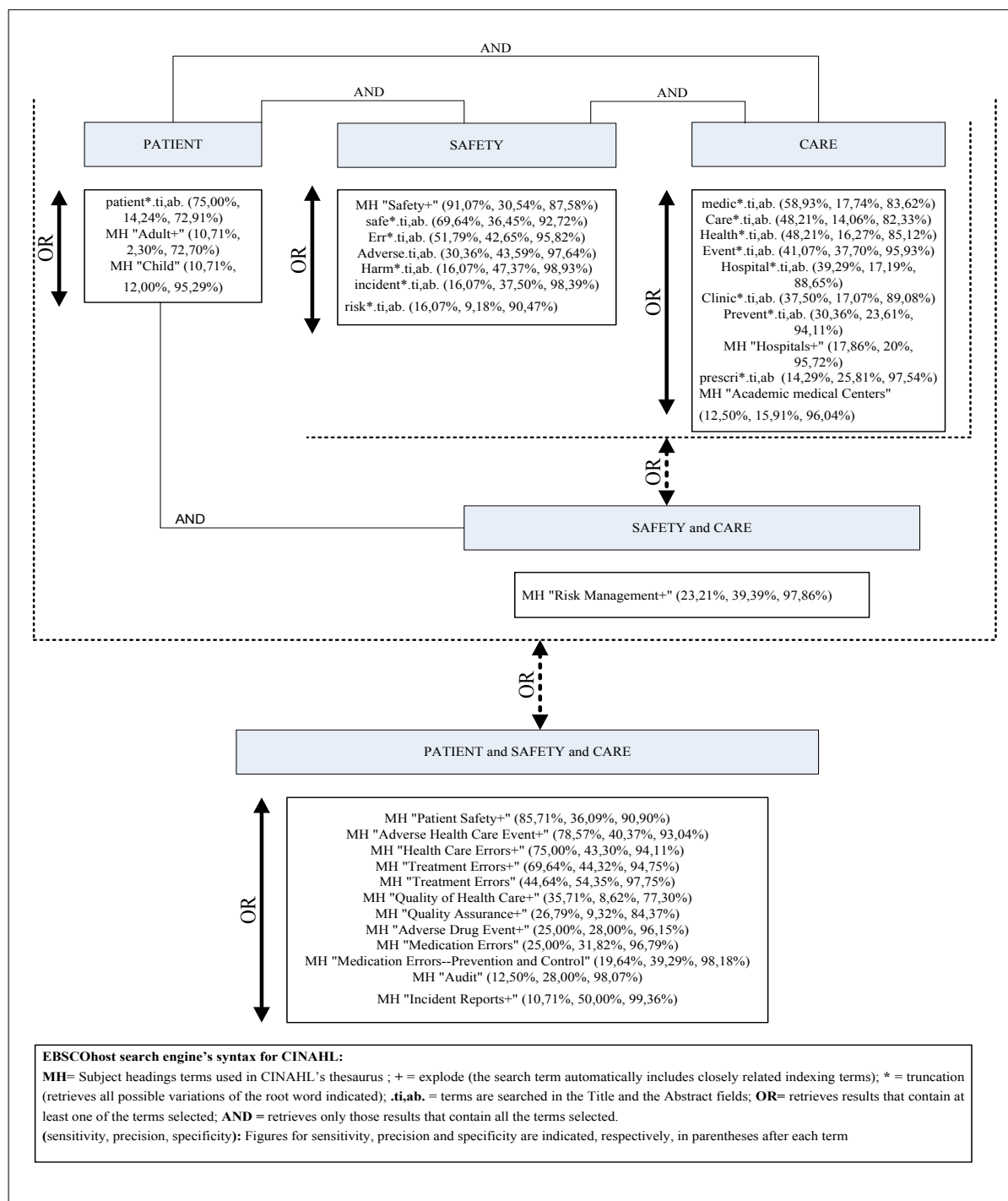


Figure 4: The conceptual design showing individual terms to be combined for final searches in CINAHL



All combinations illustrated in Figure 2, Figure 3 and Figure 4 were tested and classified according to sensitivity, precision, the product of both, and specificity. Together, sensitivity and specificity are used to assess the intrinsic value of the strategies: high sensitivity means nothing without high specificity. A strategy is considered useful if the sum of sensitivity and specificity exceeds 100% ([Straus 2005](#)).

The objective was to present three types of search strategies for three different types of users, based on Boynton et al.'s terminology ([Boynton, Glanville et al. 1998](#)) The most sensitive strategies are for “sensitivity maximizers”—users prepared to read a large number of documents for the certainty of identifying the majority of documents on the topic. The more precise strategies are aimed at “precision maximizers”—users who want to identify a maximum number of documents with minimum resource outlay. Between these extremes are strategies for users seeking a reasonable compromise between sensitivity and precision.

Table 4, Table 5 and, Table 6 present these different groups of strategies for the three databases. The performances of the search strategies in the validation gold standard are presented in parentheses. The last row presents the capacity of the filters developed using Westwood's study to identify the composition gold standard and validation gold standard articles.

Table 4: Final strategies for Medline (performances of the search strategies in the validation gold standard are presented in parentheses).

	SENSITIVITY	PRECISION	SPECIFICITY	SENSITIVITY * PRECISION	NUMBER OF ARTICLES TO READ
The 3 strategies with the greatest sensitivity					
MEDLINE STRATEGY S1 (<i>exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit/</i>)	100.00% (100%)	8.28% (3.51%)	45.56% (59.36%)	8.28% (3.51%)	12 (28)
MEDLINE STRATEGY S2 (<i>safe\$.ti.ab. OR exp Safety/ OR Err\$.ti.ab. OR Adverse.ti.ab.) AND (exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit/</i>)	98.80% (95.31%)	29.08% (17.63%)	88.17% (93.41%)	28.73% (16.80%)	3 (6)
MEDLINE STRATEGY S3 (<i>safe\$.ti.ab. OR exp Safety/ OR Err\$.ti.ab. OR Adverse.ti.ab.)</i>	98.80% (95.31%)	25.87% (13.96%)	86.09% (91.31%)	25.56% (13.30%)	4 (7)
The 3 strategies with the greatest precision					
MEDLINE STRATEGY P1 <i>*Medical Errors/</i>	45.78% (75%)	51.35% (36.64%)	97.87% (98.08%)	23.51% (27.48%)	2 (3)
MEDLINE STRATEGY P2 (<i>exp *Medical Errors/ OR *Safety Management/</i>) <i>AND (patient\$.ti.ab.)</i>	68.67% (50%)	51.35% (32%)	96.80% (98.43%)	35.27% (16%)	2 (3)
MEDLINE STRATEGY P3 (<i>exp *Risk Management/ AND</i>) <i>(patient\$.ti.ab.)</i>	60.24% (42.19%)	51.02% (34.18%)	97.16% (98.80%)	30.74% (14.42%)	2 (3)
The 3 strategies with the best combination of sensitivity and precision					
MEDLINE STRATEGY SP1 (<i>safe\$.ti.ab. OR exp Safety/ OR Err\$.ti.ab. OR Adverse.ti.ab.) AND (exp *Risk Management/ OR exp *Medical Errors/ OR *Safety Management/</i>)	92.77% (95.31%)	44.25% (32.111%)	94.26% (97.02%)	41.05% (30.60%)	2 (3)
MEDLINE STRATEGY SP2 (<i>exp *Medical Errors/ OR *Safety Management/ AND</i>) <i>(safe\$.ti.ab. OR exp Safety/ OR Err\$.ti.ab. OR Adverse.ti.ab.)</i>	89.16% (93.75%)	44.31% (32.09%)	94.50% (97.07%)	39.51% (30.08%)	2 (3)
MEDLINE STRATEGY SP3 (<i>safe\$.ti.ab. OR exp Safety/ OR Err\$.ti.ab. OR Adverse.ti.ab.) AND (exp *Risk Management/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit/</i>)	95.18% (95.31%)	41.15% (28.77%)	93.31% (96.51%)	39.16% (27.42%)	2 (3)
	VALIDATION				
STRATEGY DERIVED FROM WESTWOOD IN MEDLINE	71.08% (75.00%)	46.46% (35.29%)	95.98% (97.97%)	33.02% (26.47%)	2 (3)

Table 5: Final strategies for Embase (performances of the search strategies in the validation gold standard are presented in parentheses).

	SENSITIVITY	PRECISION	SPECIFICITY	SENSITIVITY * PRECISION	NUMBER OF ARTICLES TO READ
The 3 strategies with the greatest sensitivity					
EMBASE STRATEGY S1 (Exp Safety/ OR Safe\$.ti.ab. OR Err\$.ti.ab. OR Adverse.ti.ab) AND [(Exp Health Care Quality/ OR Risk assessment/) OR (exp *Safety/ AND *Medical Practice)]	100.00% (95.00%)	30.04% (19.52%)	86.60% (92.13%)	30.04% (18.54%)	3 (5)
EMBASE STRATEGY S2 (Exp Safety/ OR Safe\$.ti.ab. OR Err\$.ti.ab. OR Adverse.ti.ab) AND [exp Health Care Quality/ OR ((Exp Safety/ OR Err\$.ti.ab.) AND (Exp Health Care Organization/ OR Exp Health Care/ OR Medic\$.ti.ab. OR Exp Health Care System/ OR Hospital\$.ti.ab.))]	98.73% (95.00%)	30.23% (19.72%)	86.89% (92.23%)	29.85% (18.74%)	3 (5)
EMBASE STRATEGY S3 (exp Medical Error/ OR Patient Safety/ OR Exp Iatrogenic disease/) OR [(Exp Safety/ OR Safe\$.ti.ab. OR Err\$.ti.ab. OR Adverse.ti.ab.) AND [exp Health Care Quality/ OR (exp *Safety/ AND (Exp Health Care Organization/ OR Exp Health Care/ OR Medic\$.ti.ab. OR Exp Health Care System/ OR Hospital\$.ti.ab.))]]	94.94% (93.33%)	27.68% (16.67%)	85.72% (90.62%)	26.27% (15.56%)	4 (6)
The 3 strategies with the greatest precision					
EMBASE STRATEGY P1 (Exp Safety/ OR Err\$.ti.ab.) AND (patient\$.ti.ab.) AND (Exp Health Care Organization/ OR Exp Health Care/ OR Medic\$.ti.ab. OR Exp Health Care System/ OR Hospital\$.ti.ab.)	65.82% (53.33%)	45.61% (45.71%)	95.48% (98.73%)	30.02% (24.38%)	2 (2)
EMBASE STRATEGY P2 (Exp Safety/ OR Err\$.ti.ab.) AND [(Patient Care/) OR [(patient\$.ti.ab.) AND (Exp Health Care Organization/ OR Exp Health Care/ OR Medic\$.ti.ab. OR Exp Health Care System/ OR Hospital\$.ti.ab.)]]	67.09% (53.33%)	44.92% (45.71%)	95.27% (98.73%)	30.13% (24.38%)	2 (2)
EMBASE STRATEGY P3 (Exp Safety/ OR Err\$.ti.ab.) AND [(Patient Care/) OR [(patient\$.ti.ab.) AND (Human/ OR Human.ti.ab.) AND (Exp Health Care Organization/ OR Health\$.ti.ab.)]]	68.35% (55%)	43.20% (48.53%)	94.83% (98.83%)	29.53% (26.69%)	2 (2)
The 3 strategies with the best combination of sensitivity and precision					
EMBASE STRATEGY SP1 [(Exp Safety/ OR Err\$.ti.ab.) AND (Health Care Quality/) OR [(Exp Safety/ OR Safe\$.ti.ab. OR Err\$.ti.ab. OR Adverse.ti.ab.) AND (Exp Health Care Organization/ OR Health\$.ti.ab.)]	79.75% (81.67%)	42.00% (40.50%)	93.66% (97.59%)	33.49% (33.07%)	2 (2)
EMBASE STRATEGY SP2 (Exp Safety/ OR Safe\$.ti.ab. OR Err\$.ti.ab. OR Adverse.ti.ab.) AND [(Health Care Quality/ OR ((Exp Safety/ OR Err\$.ti.ab.) AND (Exp Health Care Organization/ OR Health\$.ti.ab.)]]	81.01% (85.00%)	40.51% (38.93%)	93.15% (97.32%)	32.82% (33.09%)	2 (3)
EMBASE STRATEGY SP3 (Exp Safety/ OR Safe\$.ti.ab. OR Err\$.ti.ab. OR Adverse.ti.ab.) AND [Health Care Quality OR (Exp Health Care Organization/ OR Health\$.ti.ab.)]	87.34% (88.33%)	36.51% (35.10%)	91.26% (96.72%)	31.89% (31.00%)	3 (3)
	VALIDATION				
STRATEGY DERIVED FROM WESTWOOD IN EMBASE	10.13% (8.33%)	61.54% (71.43%)	99.64% (99.93%)	6.23% (5.95%)	2 (1)

Table 6: Final strategies for CINAHL (performances of the search strategies in the validation gold standard are presented in parentheses).

	SENSITIVITY	PRECISION	SPECIFICITY	SENSITIVITY * PRECISION	NUMBER OF ARTICLES TO READ
The 3 strategies with the greatest sensitivity					
CINAHL STRATEGY S1 [(Tx Health* OR Tx Care* OR Tx Prevent*) AND (Tx Safe* OR Tx Err*) AND (MH "Child" OR Tx Patient* OR Tx Human)] OR MH "Health Care Errors+".	100.00% (100%)	18.30% (18.71%)	73.23% (89.35%)	18.30% (18.71%)	5 5
CINAHL STRATEGY S2 (Tx Health* OR Tx Care* OR Tx Prevent*) AND (Tx Safe* OR Tx Err*)	98.21% (96.15%)	17.41% (17.06%)	72.06% (88.55%)	17.09% (16.41%)	6 6
CINAHL STRATEGY S3 [(Tx Health* OR Tx Care* OR Tx Prevent*) AND (Tx Safe* OR Tx Err*)] AND (MH "Child" OR Tx Patient* OR Tx Human).	96.43% (86.54%)	18.06% (18.37%)	73.77% (90.58%)	17.42% (15.89%)	6 5
The 3 strategies with the greatest precision					
CINAHL STRATEGY P1 Tx Harm* AND Tx Event*	14.29% (9.62%)	61.54% (83.33%)	99.46% (99.95%)	8.79% (8.01%)	2 1
CINAHL STRATEGY P2 Tx Err* AND Tx Event*	39.29% (36.54%)	55.00% (59.38%)	98.07% (99.39%)	21.61% (19.28%)	2 2
CINAHL STRATEGY P3 [(Tx Harm* AND Tx Event*) OR MH "Risk Management+"] OR (MH "Child" OR Tx Patient* OR Tx Human)	35.71% (36.54%)	50.00% (52.78%)	97.86% (99.20%)	17.86% (19.28%)	2 2
The 3 strategies with the best combination of sensitivity and precision					
CINAHL STRATEGY SP1 MH "Health Care Errors+"	75.00% (90.38%)	43.30% (38.84%)	94.11% (96.51%)	32.47% (35.11%)	2 3
CINAHL STRATEGY SP2 [(Tx Err* AND (Tx Health* OR Tx Care* OR Tx Prevent*)) OR MH "Risk Management+"] AND (MH "Child" OR Tx Patient* OR Tx Human)]OR MH "Health Care Errors+".	85.71% (98.08%)	36.36% (33.77%)	91.01% (95.29%)	31.17% (33.13%)	3 3
CINAHL STRATEGY SP3 [(Tx Err* AND (Tx Health* OR Tx Care* OR Tx Prevent*)) OR MH "Risk Management+"] AND (MH "Child" OR Tx Patient* OR Tx Human)]	82.14% (84.62%)	37.10% (37.29%)	91.65% (96.51%)	30.47% (31.55%)	3 3
VALIDATION					
STRATEGY DERIVED FROM WESTWOOD IN CINAHL	(94.64%) (100.00%)	(26.37%) (19.62%)	(84.15%) (89.97%)	(24.96%) (19.62%)	(4) (5)

We should mention that none of these filters impose any restrictions on types of articles, language, age groups, subject groups, etc.

Using a conceptual design allowed us to create some interesting combinations. The strategies devised were very sensitive and highly precise. The fact that they were as effective in the validation group attested to the construction procedure's strong internal validity. The sum of sensitivity and specificity largely exceed 100% for most of the strategies.

The Westwood strategy compares very well to the most precise strategy developed for Medline. On the other hand, only two of the strategies developed had a lower sensitivity than the Westwood strategy.

The search strategies developed for Embase are systematically more sensitive than that proposed by Westwood et al., and their precision is equal to that of the most precise strategy proposed for Embase.

As for CINAHL, the strategy obtained is more precise and just as sensitive as the most sensitive strategy we developed. On the other hand, we propose other, more precise strategies that offer a better balance between sensitivity and precision.

Discussion

Search filters developed using third-generation methods offer numerous advantages over filters developed more intuitively. Nevertheless, they are not without shortcomings, some of which apply to the specific case presented in this article.

First, the method is still somewhat subjective. Some decisions were taken arbitrarily: the selection of key words to be retained from the texts for the combination of terms; the classification of the index terms in the conceptual design when the databases did not clearly define these terms; the decision to use the "explode" function, which sacrifices precision, for certain index terms; and finally, the decision to set 10% as the minimal

sensitivity level required for including individual terms in the search filters (combining the most sensitive terms does not necessarily produce more sensitive strategies).

We also did not do any preliminary study to determine the size of the gold standard. Because individual terms making up the search filters were selected based on frequency, the size of the gold standard clearly has an impact on what terms are ultimately selected.

Another potential for bias is in the choice of journals from which the gold standard is extracted. We chose mostly academic journals in medicine and nursing that publish articles in English. Clearly these are not representative of all journals that could provide material on patient safety, neither in terms of content, nor of disciplines represented or in quality of indexing.

Despite these biases, the search filters proposed in this article can lay claim to several strengths. First, an effort was made to ensure temporal representativity. The gold standard articles span from 2000 to 2006. Given the dynamic nature of index terms, this is an advantage over filters developed using articles from a single year. An effort was also made to assess the internal validity, shown to be very good. Use of the conceptual design allowed us to combine very effective individual terms logically, which ultimately resulted in filters that performed very well with regard to sensitivity, specificity and precision.

Conclusion

We offer a range of high-performance search filters to users wishing to retrieve articles on patient safety in three well-known and often-used databases: Medline, Embase and CINAHL. The choice of which strategy to use will depend on the needs of the users and as well as on their temporal, financial and human resources constraints.

Search filters are not a panacea, but used appropriately they are a valuable asset in carrying out a systematic literature review or for clinicians or researchers requiring quick searches with high precision.

Competing interests

The authors declare that they have no competing interests.

Contributors

Affaud Anaïs Tanon, François Champagne and André-Pierre Contandriopoulos contributed to the conception and design of the study. Affaud Anaïs Tanon contributed to the acquisition, analysis and interpretation of the data. Derson Bruno and Vary Jacquet contributed to the acquisition of the data. Affaud Anaïs Tanon was involved in drafting the article. All authors were involved in critically revising the article and gave final approval of the version submitted to be published.

Acknowledgements

This research was funded by the Canadian Institutes of Health Research (CIHR).

Scholarships were offered by:

- the program AnÉIS (*Analyse et Évaluation des Interventions en Santé / Analysis & Evaluation of Health Interventions*);
- the SSHRC (*Social Sciences and Humanities Research Council of Canada*);
- the GIRU (*Groupe Interuniversitaire de Recherche sur les Urgences / Interuniversity Research Group on Emergencies*);
- the Government of Côte d'Ivoire,
- the University of Montreal;
- and a research project funded by the CIHR (*the Canadian Institutes of Health Research*).

Team includes Affaud Anaïs Tanon, François Champagne, André-Pierre Contandriopoulos, Marie-Pascale Pomey, Alain Vadeboncoeur, Hung Nguyen, Derson Bruno and Vary Jacquet.

Chapitre 5 : Une méthode pour l'évaluation de la validité et de l'utilité des filtres de recherche bibliographique.

Paper 2: Retrieving information in health bibliographic databases through optimized bibliographic search filters: assessing validity and usefulness

Affaud Anaïs Tanon, François Champagne, André-Pierre Contandriopoulos

Abstract

Optimized bibliographic health filters are becoming increasingly popular in the health field as tools to effectively retrieve papers in bibliographic databases. The method used to establish their validity, which involves assessing the performance of the filters in an external gold standard, often gives confusing results, suggesting that search filters are not actually very useful. However, validity is not about yielding the same results outside the experimental environment of search filters' development. Rather, it is about making sure that the instrument measures what it is supposed to measure and that the inferences made about a specific group of subjects as a result are appropriate for a given purpose. The criterion-related validity framework, which mirrors the framework used to develop bibliographic search filters and assess their performance, offers a more suitable alternative to assess the validity of search filters' interpretation and their utility. To be effective, this framework, inspired by the psychometric field, should cohesively integrate the constraints imposed by the different types of bibliographic search filters, their specific features as measuring instruments, and the structure and features of the bibliographic databases for which they are built.

Key words: bibliographic databases, information storage and retrieval, literature review, evidence-based medicine, medical informatics

Introduction

Bibliographic search filters are logical combinations of text terms (words found in the title, abstract, or body of a document) and indexing terms (controlled vocabulary of a database) designed to identify studies meeting specific methodological criteria or covering a specific topic in a bibliographic database. They gained popularity in the health field as part of the evidence-based medicine movement ([Sackett, Straus et al. 2000](#)). They are called “optimized search filters” when more objective approaches are involved in their design, from the selection of the search terms to the evaluation of their performance using a preselected set of relevant papers, called the “gold standard” ([Jenkins 2004](#)).

As hundreds of resources are added every day to those already available in health bibliographic databases ([Hunter and Cohen 2006](#); [U.S. National Library of Medicine - National Institutes of Health 2011](#)), optimized search filters can be a very valuable tool for effectively retrieving literature indexed in bibliographic databases ([Ford 1999](#); [Kramer 2011](#)). They can be maximized for sensitivity, specificity, precision or a balance of sensitivity and precision. Sensitivity and precision are the two indicators most often used to compare search filter performance ([Hersh 2009](#)). Sensitivity is the proportion of relevant documents retrieved by a search filter, while precision is the proportion of documents retrieved by a search filter that is relevant. These two indicators move in opposite directions: the higher the sensitivity of a search filter, the lower its precision ([Buckland and Gey 1994](#)). Specificity is the complement of sensitivity. It measures the proportion of irrelevant documents that are not retrieved by a search filter. The product of sensitivity and precision is an indicator that offers a trade-off between sensitivity and precision. The number of articles to read (Number needed to read) is the inverse of precision. This indicator provides information on the number of items one needs to read before finding a relevant article retrieved by the search filter ([Bachmann, Coray et al. 2002](#)) (see Table 7).

Some authors consider the high performance of optimized search filters reported in the literature to be misleading. They argue that the usefulness of published search filters is overestimated and that, in reality, their performance is often below what is claimed ([Leeflang, Scholten et al. 2006](#); [Ritchie, Glanville et al. 2007](#); [Whiting, Westwood et al. 2011](#)). To support their argument, these authors have tested search filters published in the literature in an external gold standard composed of the same kind of papers the published filters were supposed to retrieve. The external gold standard was usually made up of studies included in systematic reviews where the filters being tested could have been used to retrieve relevant papers. They examined the percentage of papers included in the systematic reviews that were not retrieved by the filters as well as the reduction in the search effort provided by the filters. They found a very high percentage of papers missed by the filters assessed and observed that the filters did not make the retrieval task easier. They thus concluded search filters should not be used.

Other authors responded by pointing out methodological flaws in the testing; they also noted that the non-retrieved papers were not examined to determine whether they met the inclusion criteria of the filter assessed, whether they were not retrieved because of inconsistencies in indexing, or whether their non-inclusion would have affected the conclusions of the study ([Kastner, Wilczynski et al. 2009](#)). These authors cautioned that the sensitive search filters proposed in the literature should not be considered as the only tool useful for retrieval, but rather as one convenient tool that can provide rapid results as a starting point for a systematic review.

Underlying these debates are issues of validity, and more precisely, the choice of an appropriate method to establish that validity. The only method proposed so far involves assessing the filters' performance in an external gold standard ([Jenkins 2004](#); [Glanville, Bayliss et al. 2008](#)). This is based on the assumption that a valid search filter will yield sensitivity, specificity and precision that are similar to those obtained in an experimental context when used in an external gold standard. However, this assumption is not consistent with the general understanding of validity.

Validity has no single agreed-upon definition. Indeed, there are several types of validity. Some types are appropriate for discussing the design of research projects ([Munck and Verkuilen 2005](#)). Internal validity, for example, is concerned with the robustness of the relationship of one concept to another, internal to the research question under study. External validity, on the other hand, refers to the wider generalizability of the relationship uncovered between two concepts in a particular study to other research contexts ([McDonald 2005](#)). Other types of validity are more appropriate in a measurement environment where concepts are linked to observations ([Zeller 2005](#)). Validity in that context is the extent to which any measuring instrument measures what it purports to measure, rather than reflecting some other phenomenon, i.e., non-random measurement error ([Carmines and Woods 2005b](#)). Construct validity pertains to the relationships between a particular measure and other measures consistent with theoretically derived hypotheses concerning the concepts or constructs being measured. Content validity focuses on the extent to which a particular empirical measure covers a specific domain of content. Criterion-related validity concerns the relationship or correlation between a measure and some criterion variable of interest ([Carmines and Woods 2005b](#)). There are two types of criterion-related validity: predictive validity concerns a future criterion that is correlated with the relevant measure, while concurrent validity concerns a criterion that is measured alongside the relevant measure ([Carmines and Woods 2005b](#)).

The most recent thinking on validity unifies the concept: it is more appropriate to speak of several kinds of evidence to support validity than to speak of different types of validity ([Messick 1989](#); [Messick 1995](#); [American Educational Research Association, American Psychological Association et al. 1999](#); [Kane 2001](#)).

Table 7: Classification matrix for evaluating search filters' performance

		Sample of papers		Total
		Relevant (gold standard)	Irrelevant	
Search filter	Papers retrieved	a	b	(a + b)
	Papers not retrieved	c	d	(c+d)
Total		(a+c)	(b+d)	N = (a+b+c+d)

a: number of relevant papers retrieved;
b: number of irrelevant papers retrieved;
c: number of relevant papers not retrieved;
d: number of irrelevant papers not retrieved;
(a+b): set of retrieved papers;
(c+d): set of papers not retrieved;
(a+c): set of relevant papers (gold standard);
(b+d): set of irrelevant papers;
N: Total number of papers in the sample used to assess the performance of the filter

$$\text{Sensitivity (recall)} = \frac{a}{(a+c)} = \frac{\text{Number of relevant papers retrieved}}{\text{Total number of relevant papers}}$$

$$\text{Precision} = \frac{a}{(a+b)} = \frac{\text{Number of relevant papers retrieved}}{\text{Total number of papers retrieved}}$$

$$\text{Specificity} = \frac{d}{(b+d)} = \frac{\text{Number of irrelevant papers not retrieved}}{\text{Total number of irrelevant papers}}$$

$$\text{Number needed to read} = \frac{a+b}{a}$$

$$\text{Sensitivity * Precision} = \frac{a}{(a+c)} * \frac{a}{(a+b)} = \frac{a^2}{(a+c)(a+b)}$$

Search filters are measuring instruments, in that they use a numerical scale to show whether papers possess certain characteristics. The long tradition of search filters evaluation, dating back to the 1960s with the Cranfield experiments ([Järvelin 2003](#); [Tague-Sutcliffe 2009](#); [Sanderson 2010](#)), suggests that the criterion-related validity framework is suitable to assess validity in search filters because the evaluative framework used to evaluate search filters' performance mirrors the one used to assess criterion-related validity ([Nunnally and Bernstein 1994](#); [Anastasi and Urbina 1997](#)). Unlike reliability, which is the extent to which an empirical measurement yields consistent results in repeated trials ([Carmines and Woods 2005a](#)), validity is not a general attribute of measuring instruments, but is specific to a particular purpose. Indeed, a more contemporary perspective on validity in the measurement field is that validity is “the degree to which evidence and theory

support the interpretations of test scores entailed by proposed uses of tests” ([American Educational Research Association, American Psychological Association et al. 1999](#)). The implications of this definition are that:

- Validity is not a property of the instrument, but is concerned with the extent to which the inferences based on the instrument about a specific group of subjects are appropriate for a given purpose ([Furr and Bacharach 2008](#))⁴. Consequently, for a given measuring instrument, validity could be high in regard to one purpose and very low in regard to others ([Angoff 1988](#)). In other words, the fact that a search filter designed to retrieve one specific type of papers does not retrieve other types of papers is not a sign of lack of validity. However, a single measuring instrument could also display high validity for a range of different situations. Consequently, for a single measuring instrument, there should be as many validity coefficients as contexts in which criterion-related validity is to be demonstrated ([Borneman 2010](#)).
- Validity is a matter of degree; it is not an “all-or-none” issue, but rather should be conceived in terms of strong versus weak ([Furr and Bacharach 2008](#)). It has been shown that there are situations where instruments displaying weak validity can still considerably improve selection decisions ([Taylor and Russell 1939](#); [Cronbach and Gleser 1965](#)).
- Validity should be based on evidence and theory. Many kinds of scientific evidence can be used to support the validity of test instruments’ interpretation (construct validity, instrument content, consequences of use, association with other variables, responses processes) ([Furr and Bacharach 2008](#)). Depending on the nature and the type of instruments, some kinds of evidence are more appropriate than others.

This paper presents a criterion-related validity framework inspired by the psychometric tradition that could be used to assess the validity of bibliographic search filters’ interpretation as well as their utility.

⁴ Here we use the phrase “validity of the search filter” as shorthand to refer to the fact that the scores of the search filter are validly interpreted as a measure of the paper being a gold standard paper.

The criterion-related validity framework

Establishing criterion-related validity requires using an external criterion measuring the characteristics the papers are supposed to possess, to which the filter's performance could be compared ([Anastasi and Urbina 1997](#)). In that context, the filter will be the variable called predictor (**Y**) and the instrument used to decide whether a paper belongs to the gold standard will be the variable called criterion (**X**) (see Table 8).

The evidence for concurrent criterion-related validity is provided by computing a validity coefficient, which is the correlation between the predictor and the criterion. A high correlation means the papers selected through the predictor (the search filter) are more likely to belong to the gold standard than those not selected. The validity coefficient will establish with what degree of confidence the performance on the criterion (**X**) could be predicted from the performance on the predictor (**Y**).

When both the predictor and the criterion are measured on a binary scale, their correlation is measured by the phi (ϕ) coefficient, which is a variation of Pearson's correlation coefficient (r) ([Ley 1972](#)). The phi coefficient varies between -1 and 1, the negative or positive sign indicating the direction of the relationship, and the absolute value indicating the strength of the association. The phi coefficient could be interpreted as Pearson's r coefficient along the following scale:

- little to no correlation ($0.00 \leq |r| < 0.30$);
- low correlation ($0.30 \leq |r| < 0.49$);
- moderate correlation ($0.50 \leq |r| < 0.69$);
- high correlation ($0.70 \leq |r| < 0.89$);
- very high correlation ($0.90 \leq |r| \leq 1.00$) ([Whitford 2005](#)).

The phi coefficient is related to the chi-squared (χ^2) statistics, which can be used to test its significance (see Table 8) ([Kotz, Read et al. 1997](#)).

However, the phi coefficient should be interpreted with caution, because contrary to Pearson's r coefficient, the phi coefficient rarely reaches the extreme values of ± 1 ([Whitford 2005](#)). Indeed, the phi coefficient is extremely sensitive to the difference between the mean values of the predictor and the criterion. The greater that difference, the more the possibilities of achieving a value of 1 or -1 are reduced ([Guilford 1965](#)). Consequently, the mean values of the predictor and the criterion should then be examined as part of the interpretation of the phi coefficient. Their confidence intervals could be computed using formulas for binomial parameters, such as Wald's formula or the Agresti-Coull formula, which are appropriate for almost all sample sizes and parameter values, particularly when some proportions are very small ([Agresti and Coull 1998](#)).

A more objective way to interpret the correlation is to correct the phi value for the restriction of its range. Statisticians have proposed formulas to determine the maximum value (ϕ_{\max} or φ_{\max}) and the minimum value (ϕ_{\min} or φ_{\min}) the phi coefficient could take given the means of the predictor and the criterion. The phi coefficient is reported to the maximum value and the indicator $\frac{\varphi}{\varphi_{\max}}$ is used to have a better idea of the strength of the correlation between the predictor and the criterion ([Guilford 1965](#); [Peduzzi, Detre et al. 1983](#); [Davenport and El-Sanhurry 1991](#); [Warrens 2008](#)).

Another way to interpret the phi coefficient could be to compute its square. The square of the phi coefficient will then represent the proportion of variance in the criterion that is explained by the variance in the predictor. However, this statistic, because of the restriction of the range of the coefficient, may not be very useful ([Peduzzi, Detre et al. 1983](#)).

Methodological issues such as the sample size or the degree of homogeneity in the samples should also be considered, as they can influence the value of the phi coefficient ([Pedhazur and Schmelkin 1991](#); [Nunnally and Bernstein 1994](#)).

Table 8: Indicators to assess the validity of the search filter

		X: CRITERION		Total
		Relevant (gold standard)	Irrelevant	
Y	Papers retrieved	a (True positives)	b (False positives)	(a +b)
PREDICTOR	Papers not retrieved	c (False negatives)	d (True negatives)	(c+d)
Total		(a+c)	(b+d)	N = a+b+c+d
Name of the statistic	Description	Formulas		
Phi coefficient :	The measure of association between the two binary variables X and Y	$R\phi XY = \frac{ad-bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$		
$R\phi XY$				
Test of independence	H ₀ : The predictor and the criterion are statistically independent H ₁ : The predictor and the criterion are statistically dependent Assumptions: All expected frequencies are at least 1; At most 20% of the expected frequencies are less than 5.	$\chi^2 = N R\phi XY^2$ The test is performed at a α significance level. The critical value is χ^2_{α} with degrees of freedom (df) = (r-1) (c-1) where r and c are the number of rows and columns in the table. H ₀ is rejected if $\chi^2 > \chi^2_{\alpha}$ (For two binary variables, df = 1 and at a significance level of 5%, $\chi^2_{0.05} = 3,841$)		
Phimax	The maximum value the phi coefficient could take	$Phimax = \sqrt{\frac{\frac{(c+d)(a+c)}{N} + \frac{(a+b)}{N}}{\frac{(b+d)(a+b)}{N} + \frac{(a+c)}{N}}}$, if (a+b) ≥ (a+c) $Phimax = \sqrt{\frac{\frac{(b+d)(a+b)}{N} + \frac{(a+c)}{N}}{\frac{(c+d)(a+c)}{N} + \frac{(a+b)}{N}}}$, if (a+c) ≥ (a+b)		
Phimax-adjusted phi coefficient	The phi coefficient adjusted to eliminate the range restriction effect	$\frac{R\phi XY}{R\phi \max}$		

Assessing the utility of valid search filters

The evidence for concurrent criterion-related validity could also be provided by assessing the validity of the decisions taken using the predictor ([Taylor and Russell 1939](#); [Cronbach and Gleser 1965](#); [Harter 1986](#); [Anastasi and Urbina 1997](#); [Pettersen and Ebrary 2000](#)).

Base rates (*BR*) and selection ratios (*SR*) should then be considered ([Harter 1986](#)). The “base rate” is the level of successful performance on the criterion in the population. In the search filter context, that will be the proportion of gold standard papers in the sample of papers considered for the validation process. The selection ratio is the proportion of entities

selected in the population of subjects. In the search filter context, that could be the ratio of papers retrieved by the search filter to the total number of papers submitted to the search filter, if the goal is to retain all the papers retrieved by the search filter. The selection ratio could be set at another level depending on the degree of selectivity desired. As an example, if one wished to deal with only 200 papers when 2000 papers are submitted to the filter, the selection ratio will be 200/2000, or 10%.

Improvement over chance with the chi-squared test of independence

As shown in Table 7, four outcomes are possible when comparing the classification given by a search filter (or any other selection process) to the one given by the criterion:

- True positive (*a*): a paper is retrieved and belongs to the gold standard;
- False positive (*b*): a paper is retrieved and does not belong to the gold standard;
- False negative (*c*): a paper is not retrieved and does belong to the gold standard;
- True negative (*d*): a paper is not retrieved and does not belong to the gold standard.
- $P(a) = \frac{a}{N}$; $P(b) = \frac{b}{N}$; $P(c) = \frac{c}{N}$; $P(d) = \frac{d}{N}$

The chi-squared test of independence provides a useful way to measure how the use of a validated search filter can improve the selection of papers in comparison with a purely random selection process ([McCord 1980](#); [Loeber and Dishion 1983](#)).

Under the null hypothesis that the frequencies distribution of two variables presented in a 2 x 2 table are independent from each other, the general formula to compute expected frequencies of each of the four outcomes of the table is $E_{ij} = \frac{R_i * C_j}{N}$, where E_{ij} is the expected frequency for the cell in row i and column j , R_i is the total number of papers in row i , C_j is the total number of papers in column j , and N is the total number of papers in the whole table. The probability of getting the expected frequency is given by $P(E_{ij}) = \frac{E_{ij}}{N} = \frac{R_i * C_j}{N^2}$ ([Weiss and Hassett 1991](#)).

Consequently, the specific formulas for the four outcomes presented in Table 8 under the null hypothesis that the selection of the papers is independent from whether or not they belong to the gold standard, are:

$$P(E_a) = \frac{(a+b)(a+c)}{N^2}; P(E_b) = \frac{(a+b)(b+c)}{N^2}; P(E_c) = \frac{(a+c)(c+d)}{N^2}; P(E_d) = \frac{(b+d)(c+d)}{N^2}$$

The base rate (BR) = $\frac{(a+c)}{N}$ and the selection ratio (SR) = $\frac{(a+b)}{N}$ can be used to reformulate the probability of each of the four outcomes:

- $P(E_a) = BR * SR$
- $P(E_b) = SR - P(E_a)$
- $P(E_c) = BR - P(E_a)$
- $P(E_d) = 1 - P(E_a) - P(E_b) - P(E_c)$
- $P(\text{Expected Correct Decisions}) = P(E_a) + P(E_d)$

When a validated filter is used to guide the selection process, the formula for the phi coefficient ([Ley 1972](#); [Kotz, Read et al. 1997](#)) can be used to show the improvement over chance using the base rate and the selection ratio:

$$P(\text{True Positive}) = P(a) = P(E_a) + \mathbf{R\phi XY} * \sqrt{(BR(1-BR)SR(1-SR))}$$

The demonstration is as follows:

- $\mathbf{R\phi XY} = \frac{ad-bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$;
- $\mathbf{R\phi XY} = \frac{\frac{a}{N} - \left(\frac{a+b}{N}\right)\left(\frac{a+c}{N}\right)}{\sqrt{\left(\frac{a+c}{N}\right)\left(\frac{b+d}{N}\right)\left(\frac{a+b}{N}\right)\left(\frac{c+d}{N}\right)}}$;
- $\mathbf{R\phi XY} = \frac{\frac{a}{N} - BR * SR}{\sqrt{(BR)(1-BR)(SR)(1-SR)}}$;
- $\frac{a}{N} - BR * SR = \mathbf{R\phi XY} * \sqrt{(BR)(1-BR)(SR)(1-SR)}$;
- $\frac{a}{N} = BR * SR + \mathbf{R\phi XY} * \sqrt{(BR)(1-BR)(SR)(1-SR)}$;
- $P(a) = P(E_a) + \mathbf{R\phi XY} * \sqrt{(BR)(1-BR)(SR)(1-SR)}$;

The other probabilities can be obtained from the following formulas using the base rate and the selection ratio:

- $P(b) = SR - P(a)$
- $P(c) = BR - P(a)$
- $P(d) = 1 - P(a) - P(b) - P(c)$
- $P(\text{Correct Decisions}) = P(a) + P(d)$

The numerical example in Table 9 illustrates the impact of a validated search filter. In this example, with a base rate of 30% and a selection ratio of 45%, 13.5% of the random decisions are true positives, and 38.5% are true negatives, which means 52% of the random decisions are correct (true positives and true negatives) and 48% of the random decisions are incorrect (false positives and false negatives).

Using a validated search filter, with a validity coefficient of 0.6 to improve the selection of papers, results in doubling the probability of retrieving true positives. Indeed, 27.1% of the decisions made using the search filter are now true positives, and 52.1% of the decisions made using the filter are true negatives, which means 79.2% of the decisions are correct (true positives and true negatives) and 20.8% of the decisions are incorrect (false positives and false negatives).

Using this validated search filter, the percentage of correct decisions rises from 52% to 79.2%, which represents a 27.2% improvement over chance.

Table 9: A numerical example of the impact of a validated search filter, using the chi-squared test of independence.

Selection process	Papers selected Papers not selected	Set of papers		Total
		Relevant (gold standard) a = 200 c = 100 (a+c) = 300	Irrelevant b = 250 d = 450 (b+d) = 700	
Total				(a + b) = 450 (c+d) = 550 N = 1000

$$BR = \frac{300}{1000} = 30\%$$

$$SR = \frac{450}{1000} = 45\%$$

$$P(a) = \frac{450}{1000} * \frac{300}{1000} = 13.5\%$$

$$P(b) = SR - P(a) = 45\% - 13.5\% = 31.5\%$$

$$P(c) = BR - P(a) = 30\% - 13.5\% = 16.5\%$$

$$P(d) = 1 - P(a) - P(b) - P(c) = 1 - 13.5\% - 31.5\% - 16.5\% = 38.5\%$$

A filter with a validity coefficient of 0.6 is used to select the papers:

$$\begin{aligned} P(a) &= BR * SR + R\phi XY * \sqrt{(BR(1-BR)SR(1-SR))} \\ &= 13.5\% + 0.6 \sqrt{0.3(1-0.3) * 0.45(1-0.45)} \\ &= 13.5\% + 0.6 \sqrt{0.21 * 0.2475} \\ &= 13.5\% + 13.7\% \\ &= 27.2\% \end{aligned}$$

$$P(b) = 45\% - 27.1\% = 17.9\%$$

$$P(c) = 30\% - 27.1\% = 2.8\%$$

$$P(d) = 52.1\%$$

$$P(a + d) = 27.2\% + 52.1\% = 79.3\%$$

Relative improvement over chance using the RIOC statistic

When the base rate of a phenomenon is very low or very high, or when there is a very great discrepancy between the base rate and the selection ratio (the mean values of the criterion and the predictor), the correct prediction or classification of individuals in a 2 x 2 table are impaired ([Farrington and Loeber 1989](#); [Copas and Loeber 1990](#); [Cairney and Streiner 2011](#)). As outlined by [Copas and Loeber \(1990\)](#), given a study population of 1000 subjects, with a base rate of 20% and a selection ratio of 40%, the valid positive cell (the true positives) can contain no more than 200 subjects, and the valid negative cell (the true negative) can contain no more than 600 subjects. Consequently, only 800 (200 + 600) can be predicted at best. This is called the maximum ceiling bias. The larger the difference between the base rate and the selection ratio, the greater the proportion of subjects that cannot be correctly predicted.

The Relative Improvement Over Chance (RIOC) is an index of predictive association proposed by [Copas and Loeber \(1990\)](#) that combines corrections for chance and for differences in maximum values in 2 x 2 tables. It varies between 0 and 1 and offers the possibility to calculate confidence intervals and to test statistical significance between RIOCs ([Copas and Loeber 1990](#)).

The formulas for the RIOC, its confidence interval and its significance testing as given by Copas and Loeber ([Copas and Loeber 1990](#)) are presented in Table 10.

When applying these formulas to the same numerical example presented previously (See Table 11) the value of the RIOC obtained is 0.39. It is significantly different from zero and lies in a 95% confidence interval of [**0.308; 0.480**]. The RIOC value is higher than the value reported using the χ^2 test of independence, which was 0.272, suggesting that, once correction for maximum ceiling bias is taken into account, the predictive efficiency of the search filter tested appears better than previously reported. We can conclude that using the validated search filter improves paper selection over and above that achieved by chance alone by 39%.

Chance selection is not the only alternative selection strategy to which a search filter could be compared. When other strategies are available, the validity argument could include a comparison between the new filter and those already available. To determine how worthwhile the increase in the probability of making good decisions is, other information can be used, such as the cost of the information retrieval task, the opportunity cost of non-retrieved papers, or the time needed to sort the papers retrieved. Decision theory offers interesting frameworks to select comparators and to value the improvement in selection decisions and make more complete arguments about the usefulness of measuring instruments ([Cronbach and Gleser 1965](#)).

Table 10: Formulas to assess the Relative Improvement Over Chance

Selection process Total	Papers selected Papers not selected	Set of papers		Total $R_1 = (a+b)$ $R_2 = (c+d)$ $N = (a + b+ c+d)$
		Relevant (gold standard) a c $C_1 = (a+c)$	Irrelevant b d $C_2 = (b+d)$	
Statistic:		Formula:		
The Relative Improvement Over Chance	$RIOC$	$= \frac{\text{Total Correct} - \text{Chance Correct}}{\text{Maximum Correct} - \text{Chance Correct}} * 100$		
	$RIOC$	$= \frac{(a+d) - \left(\frac{(a+c)(a+b)}{N} + \frac{(b+d)(c+d)}{N} \right)}{\max(a+d) - \left(\frac{(a+c)(a+b)}{N} + \frac{(b+d)(c+d)}{N} \right)} * 100$		
	With:	$\text{Total Correct} = (a + d)$ $\text{Chance Correct} = \left(\frac{(a+c)(a+b)}{N} + \frac{(b+d)(c+d)}{N} \right)$ $\text{Max}(a) = \text{Min}((a + c) (a + b))$ $\text{Max}(d) = \text{Min}((b + d) (c + d))$ $\text{Maximum Correct} = \text{Max}(a) + \text{Max}(d)$		
The Confidence Interval formula for the RIOC	$RIOC \pm S_{RIOC}$	where:		
	$S_{RIOC} = \sqrt{S_{RIOC}^2}$	and $S_{RIOC}^2 = \frac{Nc(NC_1R_2 + c(NR_1 + R_1C_1 - 2NC_1 - N^2) + 2Nc^2)}{R_2^3C_1^3}$		
		$R_1 = (a + b); R_2 = (c + d); C_1 = (a + c); C_2 = (b + d)$		
	$(RIOC S_{RIOC}^*)$	$(RIOC S_{RIOC}^*) = \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$		
		To test the difference between the RIOC and the chance value of zero, the quantity $(RIOC S_{RIOC}^*)$ is tested as χ^2 on 1 d.f.		
Statistics for the comparison of two RIOCs		$\frac{RIOC_1 - RIOC_2}{\sqrt{(S_{RIOC1}^2 + S_{RIOC2}^2)}}$		
		There is a significant difference between $RIOC_1$ et $RIOC_2$ at the 5% significance level if the statistics is greater than 2 or less than -2.		

Table 11: A numerical example of the impact of a validated search filter using the RIOC statistic

Selection process	Papers selected Papers not selected	Set of papers		Total
		Relevant (gold standard) a = 200 c = 100 C ₁ = (a+c) = 300	Irrelevant b = 250 d = 450 C ₂ = (b+d) = 700	
Total				R ₁ = (a + b) = 450 R ₂ = (c+d) = 550 N = 1000
<hr/>				
<i>Total Correct</i>	= 650			
<i>Chance Correct</i>	= $\left(\frac{(a+c)(a+b)}{N} + \frac{(b+d)(c+d)}{N} \right)$			
	= $\left(\frac{(300)(450)}{1000} + \frac{(700)(550)}{1000} \right)$			
	= 135 + 385 = 520			
<i>Max (a)</i>	= 300			
<i>Max (d)</i>	= 550			
<i>Maximum Correct</i>	= <i>Max(a)</i> + <i>Max(d)</i> = 300 + 550 = 850			
<i>RIOC</i>	= $\frac{\text{Total Correct} - \text{Chance Correct}}{\text{Maximum Correct} - \text{Chance Correct}} * 100$			
	= $\frac{650 - 520}{850 - 520} * 100 = 39\%$			
S_{RIOC}^2	= $\frac{Nc(NC_1R_2 + c(NR_1 + R_1C_1 - 2NC_1 - N^2) + 2Nc^2)}{R_2^3C_1^3} = 0.0019$			
$(RIOC S_{RIOC}^*)$	= $\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} = 81$			

Criterion-related validity of search filters: preconditions

In general, as a preliminary step before assessing criterion-related validity, it is important to demonstrate that both the predictor and the criterion are reliable instruments and possess good content validity ([Harter 1986](#); [Anastasi and Urbina 1997](#); [Pettersen and Ebrary 2000](#)).

There are several methods to assess reliability ([Carmines and Woods 2005a](#)). The **test-retest method** consists in measuring the **same** measures in the **same** population at different points in time. Reliability is measured by the correlation of the scores on the same measuring instrument at two points in time. The **alternative form method** consists in submitting the **same** population to two alternative forms of the **same** instrument. Reliability is the correlation between the alternative forms of the instrument. In the **split halves**

method, the set of items forming the measuring instrument are divided into halves, submitted to the same population. Reliability is measured by the correlation between the halves. **Internal consistency methods**, such as Cronbach's alpha, could also be used to assess reliability. Reliability is a concern when random error could affect the measurement process. This is not the case for bibliographic search filters, as these measuring instruments are administered in a computerized environment where random errors no longer exist. The only fluctuations possible would be related to systematic changes, such as modifications in the indexation process. Consequently, the test-retest method will consistently yield the same results when these systematic changes are controlled. Other methods of reliability assessment involving the analysis of sub-items cannot be applied, because bibliographic search filters are single-item instruments. They may include several search terms, but the logical operators used to combine them make them an unbreakable item.

For the criterion, however, assessing reliability may be necessary, because human judges are often involved in rating the sample of papers used to build the gold standard. Of the many methods available to assess reliability, one that could be very useful in the context of gold standard building would be to provide inter-rater reliability coefficients ([DeVellis 2005](#)).

The importance of content validity depends on the purpose for which the measuring instruments were built ([Harter 1986](#)). Content validity is required for measuring instruments used to make descriptive inferences, i.e., to infer what is really measured by the results ([Haynes, Richard et al. 1995](#)). It is of secondary importance (and sometimes not even necessary) for instruments whose objective is to make relational inferences, that is, to identify and classify items into groups using a variable that is not directly measured ([Pettersen and Ebrary 2000](#)). Based on most of the definitions found in the literature ([Jenkins 2004](#)), we defined search filters as selection instruments whose purpose is to identify papers that meet certain predetermined criteria or, to use medical language, as screening tests.

Defining bibliographic search filters in that way determines the population to which the test should be administered. Indeed, screening tests are typically administered to a

healthy population to identify those who might suffer from a particular condition ([Raffle and Muir Gray 2007](#)). Likewise, search filters designed as screening instruments are administered to all papers in a particular database to identify the relevant ones. Another practical consequence of this choice is that the content validity of the test becomes a secondary issue ([Pettersen and Ebrary 2000](#)). Depending on the quality of the indexing process in a given database, and the level of reporting consistency in the papers, the best performing terms used to form the filter may not be easily inferred by the content domain, but rather by the indexing features of the bibliographic database considered. However, search filters can serve many other purposes. As an example, search filters have already been described in the literature as diagnostic tests ([Haynes and Wilczynski 2004](#); [McKibbon, Lokker et al. 2012](#)). The purpose of diagnostic tests is to confirm or determine the presence of a disease in individuals presenting symptoms. They usually involve more sophisticated types of information ([Knottnerus and Buntinx 2009](#)). Diagnostic test scores have meanings different from those obtained in screening tests; they allow descriptive inferences about the individual being tested. Content validity of the predictor plays a larger role in this context, because the diagnostic value of a test is related to its ability to represent the domain of interest ([Anastasi and Urbina 1997](#)). The population tested is one that has already been tested using some screening instrument ([Knottnerus and Buntinx 2009](#)). The ability to develop diagnostic search filters may depend on the structure and features offered by the bibliographic database considered. Most bibliographic databases of a certain size are organized using an inverted index and Boolean logic to support the construction of search filters ([Salton and McGill 1983](#); [Manning, Raghavan et al. 2008](#)), but new interfaces integrating the latest advances in the information retrieval field offer user-centered capabilities that could be used to develop more sophisticated search filters ([Kim and Rebholz-Schuhmann 2008](#)).

On the other hand, it is very important to demonstrate that the instrument used to build the gold standard possesses good content validity, because the criterion is the most important aspect of criterion-related validity and the validity of the predictor is judged by its correlation to the criterion. It is even more important because the search terms included in the filters derive from the gold standard papers and the way they have been indexed in

the bibliographic databases. Consequently, the content validity of the instrument used to build the gold standard has a direct influence on the performance of the filters. The instrument used to represent the criterion should measure fundamental aspects of the conceptual criterion ([Harter 1986](#)). However, this is not always an easy task, especially when the conceptual criterion is not well known or is evolving. One way to demonstrate content validity could be to design an assessment tool for the instrument measuring the criterion that would be submitted to a panel of experts ([McGartland Rubio 2005](#)).

Conclusion

Validity is a very important feature of any measuring instrument and particularly of search filters. For decades, a framework involving a gold standard has been used to assess the performance of bibliographic search filters. The criterion-related validity framework mirrors that framework and can be used to assess the validity of search filters' interpretation and their utility.

Simply assessing the performance of published search filters in an external gold standard is not the appropriate way to demonstrate validity, as validity is not about yielding the same results outside the experimental environment of search filters' development, but rather, is about making sure the instrument measures what it is supposed to measure and that the inferences made based on the instrument about a specific group of subjects are appropriate for a given purpose. However, when there are strong reasons to believe a given bibliographic filter should perform very well against a particular set of papers, encountering negative or mitigated results may be indicative of indexing problems in the databases, the use of different selection criteria or of different relevance assessment criteria, etc. Analyzing these problems might reveal new contexts for which the search filters, otherwise valid in other circumstances, should be adapted. Indeed, the proposed framework is suitable for what is called a local validation study (one predictor, one criterion). However, criterion-related validity can be generalized to other predictors (comparing different filters), other criteria (comparing different criteria pertinent to the situation), other situations (comparing

different types of uses), other samples (comparing different sets of papers), and other methods (using different methods to estimate the validity coefficient) ([Harter 1986](#)). This ongoing process of “validity argument” building is particularly important when an instrument is young, because any work that clarifies possibilities, anomalies, and boundary conditions is worthwhile ([Cronbach 1988](#)) and because a criterion considered essential can evolve over time ([Harter 1986](#)).

Whatever the purpose for which search filters have been developed, it is important to note that their nature as measuring instruments differs from that of other measuring instruments generally dealt with in psychometrics or test theory. We already mentioned that, as these are informatics-related measuring tools administered to a population of inanimate subjects in a computerized environment, their reliability is not really an issue. What could be an issue are the relevance judgments involved in the validity assessment process: search filters’ developers decide what papers should be considered as gold standard based on their knowledge of the domain; the information retrieval system measures relevance objectively as a relation between the terms included in the search filters and those used to index the papers in the database; and users of the filters judge relevance based on the content of the papers, the innovation compared to what they already know, and the pertinence or situational usefulness of the content ([Lancaster and Gale 2003](#); [Saracevic 2007](#); [Cosijn 2009](#)). Relevance has played an important role in the information retrieval field in the design and evaluation of search filters. It could be interesting to assess the impact of relevance on the evaluation of search filters’ validity as well as on the assessment of their utility. Another possible challenge could be the capabilities offered by the bibliographic databases considered. For future work, the constraints imposed by the different types of bibliographic search filters, their specific features as measuring instruments, and the structure and features offered by the bibliographic databases for which they are built should be integrated cohesively into the framework used to assess the validity of search filters’ interpretation and their utility.

Competing interests

The authors declare that they have no competing interests.

Contributors

Affaud Anaïs Tanon, François Champagne, and André-Pierre Contandriopoulos were all involved in generating the ideas presented in this paper. Affaud Anaïs Tanon developed the ideas, organized the arguments, and drafted the article. All authors were involved in critically reviewing the article and gave final approval of the version submitted for publication.

Acknowledgments

This research was funded by the Canadian Institutes of Health Research (CIHR). Scholarships were provided by: the AnÉIS program (Analyse et Évaluation des Interventions en Santé / Analysis and Evaluation of Health Interventions); SSHRC (Social Sciences and Humanities Research Council of Canada); GIRU (Groupe Interuniversitaire de Recherche sur les Urgences / Interuniversity Research Group on Emergency Services); the Government of Côte d'Ivoire; the University of Montreal; and a research project funded by CIHR (Canadian Institutes of Health Research).

The project was conducted with the support of the Takemi Program in International Health at Harvard School of Public Health. We would like to thank Professors Michelle Mello and Ashish Jha, advisers to Affaud Anaïs Tanon at Harvard University. We also thank Su Golder at the University of York and Professor Kaveh Shojania at the University of Toronto for their comments on previous drafts.

Chapitre 6: La validation des filtres de recherche proposés pour Medline, Embase et CINAHL.

Paper 3: Patient safety and systematic reviews: A criterion-related validity framework to assess the validity of optimized bibliographic search filters designed for Medline, Embase, and CINAHL

Affaud Anaïs Tanon, François Champagne, André-Pierre Contandriopoulos

Abstract

Objective: To investigate the criterion-related validity of optimized bibliographic search filters for predicting whether papers indexed in Medline, Embase and CINAHL cover the subject of patient safety in health care organizations.

Methods: Data were collected using the same samples used to develop and assess the performance of the search filters in the three databases. Concurrent validity was assessed by computing phi coefficients.

Results: The phi coefficients were reported at their maximal values to compensate for range restriction and facilitate interpretation. In both samples and in all the databases, concurrent validity with the criterion was very high for the sensitive strategies as well as for the Medline balanced strategies ($\text{phi}/\text{phimax} \geq 0.90$). In both samples, it was high for most of the Embase and CINAHL balanced strategies ($0.70 \leq \text{phi}/\text{phimax} < 0.89$) and moderate for the Embase precise strategies ($0.50 \leq \text{phi}/\text{phimax} < 0.69$). It was low or moderate for most of the Medline and CINAHL precise strategies depending on the sample considered ($0.30 \leq \text{phi}/\text{phimax} < 0.69$).

Conclusion: To our knowledge, this study provides the first evidence reporting the concurrent validity of subject-oriented bibliographic search filters. The most sensitive and balanced strategies can be validly interpreted as a measure of the papers being retrieved that deal with patient safety in health care organizations. Based on these results, the authors encourage more research on the appropriate sampling method for search filters construction and validation and the use of reliable and valid instruments for gold standard selection. The more sensitive strategies should be considered first when trying to apply the filters outside the experimental context of their construction.

Keywords: bibliographic databases, information storage and retrieval, literature review, evidence-based medicine, criterion-related validity.

Introduction

Optimized bibliographic search filters are logical combinations of text terms (words found in the title, abstract, or body of a document) and indexing terms (controlled vocabulary of a database) designed to identify studies meeting specific methodological criteria or covering a specific topic in a bibliographic database. They are called “optimized search filters” when more objective approaches are involved in their design, from the selection of the search terms to the evaluation of their performance using a preselected set of relevant papers called a “gold standard” ([Jenkins 2004](#)). They can be an invaluable tool for effectively retrieving literature indexed in bibliographic databases ([Haynes, Wilczynski et al. 1994](#); [Wilczynski, Haynes et al. 2004](#); [Wong, Wilczynski et al. 2006](#)). However, the ultimate test of search filters’ utility lies in evaluating their validity to know how useful they could be outside the experimental environment in which they have been designed ([Jenkins 2004](#)).

The extent of validation varies from one study to another; this is due to the lack of a standard definition of “validation” in search filter methodology ([Jenkins 2004](#)). The method usually recommended, which consists in assessing the performance of the filters in an external gold standard ([Glanville, Bayliss et al. 2008](#)), often gives confusing results.

Depending on the external gold standard chosen for the validity study, the performance of the filters under scrutiny is either confirmed ([Kastner, Wilczynski et al. 2009](#)) or contradicted ([Leeflang, Scholten et al. 2006](#)), suggesting that, in reality, search filters are not very useful.

Bibliographic search filters can be defined as measuring instruments, in that they are designed to measure the extent to which a paper indexed in a particular database could fulfill the information needs of a particular user; this is done through a quantitative scale (mostly dichotomous) used to give scores to papers indexed in a bibliographic database ([Manning, Raghavan et al. 2008](#)). The scores are assigned in a way that distinguishes papers that are relevant to a user from those that are not.

Validity in a measurement context is concerned with the extent to which the interpretations made when using a measuring instrument are appropriate for a given purpose ([Carmines and Woods 2005b](#); [Furr and Bacharach 2008](#)). It is a property of inferences, not an intrinsic property of the measuring instrument or its scores ([Messick 1989](#); [Messick 1995](#); [American Educational Research Association, American Psychological Association et al. 1999](#)). Different types of evidence could be produced to establish the validity of a measuring instrument. Traditionally, three types of validity have been considered the most important: content validity, criterion-related validity, and construct validity ([Goodwin 2002](#)). Content validity documents the extent to which the measuring instrument is representative of a certain content domain ([McGartland Rubio 2005](#)). Criterion-related validity assesses the relationship or correlation between a measure and some criterion variable of interest ([Carmines and Woods 2005b](#)); the criterion variable is supposed to be a direct measure of the concept under study ([McDonald 2005](#)). Construct validity pertains to the relationships between a particular measure and other measures consistent with theoretically derived hypotheses concerning the concepts or constructs being measured ([Carmines and Woods 2005b](#)). The most recent thinking on validity unifies the concept: it is more appropriate to speak of several kinds of evidence to support validity than to speak of different types of validity ([Messick 1989](#); [Messick 1995](#); [American Educational Research Association, American Psychological Association et al. 1999](#); [Kane 2001](#)). Evidence for validity could be based on test content, on response processes, on the instrument's internal structure, on the

relation to other variables, or on the consequence of testing ([American Educational Research Association, American Psychological Association et al. 1999](#)), see Table 12. The choice of the most appropriate type of evidence needed to establish validity will depend on considerations such as the type of measuring instrument considered, the purpose for which the measuring instrument has been built, and the type of evidence derived from the test scores ([American Educational Research Association, American Psychological Association et al. 1999](#)).

For more than four decades, bibliographic search filters have been developed and their performance has been assessed using an evaluative framework that involves a set of documents, a set of criteria to assess the documents' relevance in relation to some specific purpose, and a gold standard—a set of relevant references within the aforementioned set of documents ([Järvelin 2003](#); [Tague-Sutcliffe 2009](#); [Sanderson 2010](#)). That framework mirrors the criterion-related validity framework, in which the results obtained through a measuring instrument are compared to those obtained using a criterion variable (validity evidence based on the relation to other variables) ([Nunnally and Bernstein 1994](#); [Anastasi and Urbina 1997](#)). When people use a bibliographic search filter as a selection device, their main concern is not whether the papers retrieved using the filter possess the exact combination of search terms exhibited in the filter. Rather, they care about whether the papers have the potential to fulfill their information need. The criterion-related validity framework offers a suitable alternative to assess the validity of search filters' interpretation, especially because in criterion-related validity, it is the criterion that is of concern to the one using the measuring instrument ([Cronbach and Meehl 1955](#)).

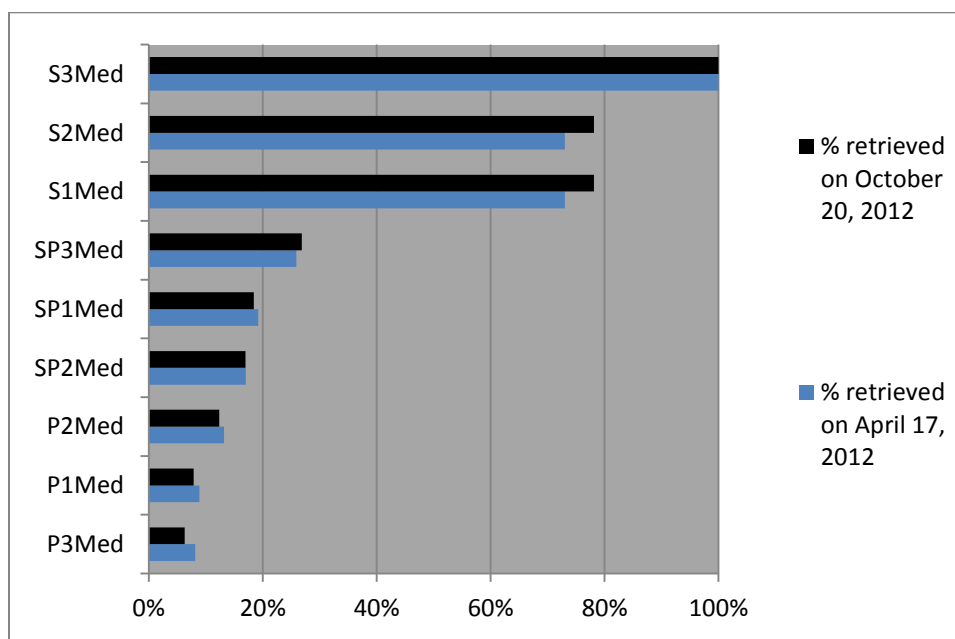
In a previous paper, we proposed 27 optimized bibliographic search filters developed between April and December 2007 to retrieve patient safety papers in Medline, Embase, and CINAHL. Two convenient samples of papers were selected in order to have two distinct gold standards that will be used to develop the filters and assess their performance ([Tanon, Champagne et al. 2010](#)).

The filters we proposed achieved very good performance in both gold standards. For all three databases, we proposed highly sensitive search filters with sensitivities in the

range of 95%–100%, precise search filters with precision in the range of 40%–60%, and balanced search filters with the product of sensitivity and precision being in the range of 30%–40% (see Annexe 5). The filters were also very specific and outperformed those found in the literature at the time ([Tanon, Champagne et al. 2010](#)). Since those filters were developed, a new indexing term, “patient safety,” has been introduced in Medline. Its scope has been defined as “efforts to reduce risk, to address and reduce incidents and accidents that may negatively impact healthcare consumers” and it is used for “patient safety”, “patient safeties”, “safety patient” and “safeties patient” ([Schulman 2011](#)).

The papers indexed under this MeSH (Medical Subject Heading) are retrieved by one of the Medline filters (S2Med) at a 100% rate. This is because, in the MeSH Tree Structure, the MeSH “patient safety” comes under the MeSH “safety”, which is “exploded” in S3Med. The papers indexed under “patient safety” are retrieved by two other filters (S1Med, S3Med) at a 70% rate (Figure 5). No other indexing vocabulary changes affecting the search filters have been identified in the other databases.

Figure 5. Percentage of the papers indexed under the MeSH “patient safety” retrieved by the filters developed for Medline



As a first step in demonstrating the usefulness of the filters outside the experimental context in which they have been developed, it is important to assess their validity.

Table 12 : Modern views on validity evidence

Type of validity evidence	Examples of validity activities
A Evidence based on test content (content-related evidence)	<ol style="list-style-type: none"> 1 Logical analysis of extent to which the test content represents the content domain. 2 Experts' evaluations of the extent to which the items or subparts of a test match the definition of the construct and/or the purpose of the test. 3 Experts' evaluations of the relevance, importance, and clarity of items or tasks. 4 Experts' evaluations of the extent to which construct underrepresentation or construct irrelevant components may give unfair advantages to one or more subgroups of examinees.
B Evidence based on response processes (construct-related evidence)	<ol style="list-style-type: none"> 1 Analyses of individual responses, via interviews with test takers. 2 Monitoring changes in, or development of, responses to performance tasks. 3 Process studies, examining similarities and differences in responses given by members of distinct subgroups of examinees. 4 Studies on the ways in which judges, observers, and interviewers collect, record, and interpret data.
C Evidence based on internal structure (construct-related evidence)	<ol style="list-style-type: none"> 1 Factor analytical studies. 2 Item analyses to examine item relationships. 3 Differential item functioning (DIF) studies.
D Evidence based on relations to other variables (D1-2: criterion-related evidence; also, concurrent and predictive validity) (D3-6: construct-related evidence)	<ol style="list-style-type: none"> 1 Correlational studies of the type and extent of relationships between scores and external variables. 2 Correlational studies of the extent to which scores forecast or predict criterion performance or scores on measures obtained at a later date. 3 Convergent validity studies, investigating the relationships between scores and other tests intended to measure similar constructs. 4 Discriminant validity studies, investigating the relationships between scores and other measures of purportedly different constructs. 5 Experimental studies, intended to test hypotheses about effects of specific interventions on test scores. 6 Known-group comparison studies, intended to test hypotheses about expected differences in test scores across specific groups of examinees.
(D7-9: criterion-related evidence)	<ol style="list-style-type: none"> 7 Studies of the effectiveness of selection, placement, and classification decisions. 8 Differential group prediction or relationship studies. 9 Studies of validity generalization.
E Evidence based on consequences of testing	<ol style="list-style-type: none"> 1 Descriptive studies of the extent to which anticipated benefits of testing are realized. 2 Descriptive studies of the extent to which unanticipated negative consequences occur.

Source: Goodwin, L.D. (2002). The types of evidence are those presented in the 1999 Standards for Educational and Psychological Testing published by the American Educational Research Association, the American Psychological Association, the National Council on Measurement in Education and the US Joint Committee on Standards for Educational and Psychological Testing. Corresponding types of validity as discussed in the 1985 Standards are given by Goodwin in parentheses.

Objective

The present paper aims to apply a criterion-related validity framework inspired by the psychometric tradition to the 27 patient safety filters developed for Medline, Embase, and CINAHL in order to assess the validity of their interpretation.

Method

The criterion (X)

The optimized bibliographic search filters were designed within the larger framework of a quantitative conceptual analysis aimed at better understanding the concept of “patient safety” and its links with the concept of “quality”, using the scientific literature on patient safety published since 1999.

With respect to that goal, it was important to build search filters capable of identifying papers in which authors explicitly defined their work as relevant to the patient safety field (as opposed to papers rated as relevant to the field by external experts) so that it would be easier to analyze the “core” of the concept and the different themes that have been added to that core over the years. Therefore, the first criterion used to select the gold standard papers was that the title, abstract or indexing terms of the papers explicitly mention classical vocabulary terms of the field such as “patient safety”, “medical errors”, “medication errors”, “adverse or undesirable events, effects or outcomes”, “medical injuries,” etc. The list of candidate terms was developed using general and specific dictionaries like the *Grand dictionnaire terminologique* ([Office Québécois de la langue française 2000](#)) and the *Canadian Patient Safety Dictionary* ([Davies, Hebert et al. 2003](#)).

As patient safety studies focused initially on hospitals and other health care organizations before extending their scope to other settings ([Elder and Dovey 2002](#); [Hammons, Piland et al. 2003](#)), we hypothesized that the discourse on the concept would be more fully articulated in the first two settings. Therefore, as a second criterion we chose papers describing patient safety in health care organizations in the context of a health care process and did not include papers dealing with subjects such as health care

services dispensed at home, the initial steps of the emergency care continuum (911 calls, care received on a scene of an illness or trauma or during transportation by ambulance, and air medical service), patient safety in public health, patient aggression and violence, patient suicide, environmental safety, safety of health care workers, and the like.

The criterion variable (**X**) received the value 1 if the paper examined dealt with patient safety in health care organizations and its authors explicitly situated their work in the patient safety field by the use of common patient safety terms; otherwise, it received the value 0.

The different steps followed to select the gold standard papers using these criteria have been described in a previous paper ([Tanon, Champagne et al. 2010](#)), and the forms used by the reviewers at each step are presented in Annexe 6. The number of papers defined as gold standard papers according to our selection criteria in each sample (GS_1 and GS_2) differs in each database due to incomplete indexation (see Table 14).

The samples

The publication source and the publication year of the gold standard papers were used to allocate the journals initially used to select the gold standard papers into two groups corresponding to each of the gold standards. The two samples of papers were composed of all the papers published in these two groups of journals ([Tanon, Champagne et al. 2010](#)). The first sample (Sample 1) was composed of papers published in JAMA (2001), Quality & Safety in Health Care (2004, 2005 and 2006) and Nursing Management (2005). The second sample (Sample 2) was composed of papers published in BMJ (2000), the New England Journal of Medicine (2003), Quality & Safety in Health Care (2002 and 2003) and Nursing Management (2006). As was the case for the gold standard papers, the size of the samples (N_1 and N_2) differs in each database due to incomplete indexation ([Tanon, Champagne et al. 2010](#)).

The proportions of gold standard papers in the two samples ($P_{X1} = \frac{GS_1}{N_1}$ and $P_{X2} = \frac{GS_2}{N_2}$) were computed, and their confidence intervals were calculated using the Agresti–Coull formula ([Agresti and Coull 1998](#)).

A chi-squared test was performed to assess the probability that the differences in the proportions of gold standard papers between the two samples were significant ([Thibault, Bagel et al. 2001](#)). The relative risk ratio $\frac{P_{X1}}{P_{X2}}$ was computed to give an idea of the likelihood of having a gold standard paper in one sample relative to the likelihood of having a gold standard paper in the other sample ([Sistrom and Garvan 2004](#)).

The predictors (Y)

The 27 search filters developed were the predictors. They were developed using the gold standard in Sample 1. For each of them we compute the proportion of papers retrieved in each sample and its confidence interval. A chi-squared test was performed to assess the probability that the differences in the proportions of papers retrieved by the filters between the two samples were significant. We also computed the relative risk ratio $\frac{P_{Y1}}{P_{Y2}}$ as an indicator of the likelihood of a paper being retrieved in the first sample relative to the likelihood of a paper being retrieved in the second sample.

The criterion-related validity framework

The supporting arguments for the use of the criterion-related validity framework as an appropriate way to assess the validity of optimized bibliographic search filters have been presented in chapter 5. The phi coefficient, a variation of Pearson's r correlation coefficient for binomial variables, was computed as the validity coefficient to assess the correlation between each of the 27 predictors and the criterion in both samples. The phi coefficient was interpreted using the ratio of its reported value to the maximal value it could take, given the restriction of its range, because we had observed important

discrepancies between the mean values of the predictors and the criterion in each sample.

The following scale was used:

- little to no correlation ($0.00 \leq \left| \frac{\varphi}{\varphi_{max}} \right| < 0.30$);
- low correlation ($0.30 \leq \left| \frac{\varphi}{\varphi_{max}} \right| < 0.49$);
- moderate correlation ($0.50 \leq \left| \frac{\varphi}{\varphi_{max}} \right| < 0.69$);
- high correlation ($0.70 \leq \left| \frac{\varphi}{\varphi_{max}} \right| < 0.89$);
- very high correlation ($0.90 \leq \left| \frac{\varphi}{\varphi_{max}} \right| \leq 1.00$).

A chi-squared (χ^2) statistic was computed to test the significance of the phi coefficient (see Table 13).

Table 13 : Indicators to assess the validity of the search filter

		X: CRITERION		Total
		Relevant (gold standard)	Irrelevant	
Y PREDICTOR	Papers retrieved	a (True positives)	b (False positives)	(a + b)
	Papers not retrieved	c (False negatives)	d (True negatives)	(c + d)
Total		(a+c)	(b+d)	N = a+b+c+d

Statistic name	Description	Formulas
Phi coefficient : $R\phi XY$	The measure of association between the two binary variables X and Y	$R\phi XY = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$
Test of independence	H ₀ : The predictor and the criterion are statistically independent H ₁ : The predictor and the criterion are statistically dependent	$\chi^2 = N R\phi XY^2$
	Assumptions: All expected frequencies are at least 1; At most 20% of the expected frequencies are less than 5.	The test is performed at a α significance level. The critical value is χ^2_{α} with degrees of freedom (df) = (r-1) (c-1) where r and c are the number of rows and columns in the table.
Phimax	The maximum value the phi coefficient could take	H ₀ is rejected if $\chi^2 > \chi^2_{\alpha}$ (For two binary variables, df = 1 and at a significance level of 5%, $\chi^2_{0.05} = 3,841$)
		$\text{Phi max} = \frac{\sqrt{\frac{(c+d)}{N} \cdot \frac{(a+c)}{N}}}{\sqrt{\frac{(b+d)}{N} \cdot \frac{(a+b)}{N}}}, \text{ if } (a+b) \geq (a+c)$
		$\text{Phi max} = \frac{\sqrt{\frac{(b+d)}{N} \cdot \frac{(a+b)}{N}}}{\sqrt{\frac{(c+d)}{N} \cdot \frac{(a+c)}{N}}}, \text{ if } (a+c) \geq (a+b)$
Phimax-adjusted phi coefficient	The phi coefficient adjusted to eliminate the range restriction effect	$\frac{R\phi XY}{R\phi \text{ max}}$

Comparing the two samples

Table 14 presents an analysis of the two samples in the three databases. On average 5% of the papers in Sample 1 were defined as gold standard papers compared to an average of 2% in Sample 2. A chi-squared test confirmed that, based on the proportion of papers defined as gold standard according to our criteria, the two samples can be considered as coming from two different populations of papers in each of the three databases (the value of the χ^2 statistics exceeds the critical value for df = 1 and $(1 - \alpha) = 95\%$, which is 1.96). Based on the value of the relative risk ratio $\frac{P_{X1}}{P_{X2}}$, Sample 1 tends to have, on average, 2.5 times more gold standard papers than Sample 2.

Table 14: Analysis of the samples used to assess criterion-related validity

	N_1	GS_1 ($a+c$) ₁	Sample 1		N_2	GS_2 ($a+c$) ₂	Sample 2		χ^2 $\alpha = 5\%$ $df = 1$	$\frac{P_{X1}}{P_{X2}}$
			P_{X1}	95% CI			P_{X2}	95% CI		
Medline	1773	83	0.047	[0.038; 0.058]	4392	64	0.015	[0.011; 0.019]	56.41	3.21
Embase	1452	79	0.054	[0.044; 0.067]	2146	60	0.020	[0.015; 0.025]	39.56	2.76
CINAHL	990	56	0.057	[0.044; 0.073]	2175	52	0.024	[0.018; 0.031]	22.01	2.37

Table 15 shows that the filters are more effective in Sample 1, the sample used to develop the filters, compared to Sample 2 and that the difference is significant. However, the ratio $\frac{P_{Y1}}{P_{Y2}}$ shows that the magnitude of that difference varies considerably depending on the filter considered, which could mean some search terms are more representative of Sample 1 than of Sample 2. The precise strategies are on average more effective in Sample 1.

From the analysis of the criterion and the predictors, we can tell the two samples come from two different populations. Therefore, the scenario considered here is a criterion-related validity generalization framework (one criterion, two different populations) rather than a double validation framework, which requires the two samples to come from the same population with respect to the criterion ([Harter 1986](#); [Pettersen and Ebrary 2000](#)).

The correlation

Generally speaking, there was an important difference between the proportions of gold standard papers in the sample (P_X) and the proportions of papers retrieved by the filters (P_Y). The difference was less important for the precise (P) strategies. However, as the phi coefficient reaches the extreme values of ± 1 only when the mean values of the criterion and the predictor equal 0.50 ([Guilford 1965](#)), the phi coefficient was adjusted to its maximum value before being interpreted.

Results

Table 16 shows that the phi coefficient is inferior to 0.30 for S1Med and P1Cin, and between 0.30 and 0.70 for all the other filters in both Sample 1 and Sample 2. These coefficients suggest a weak validity for S1Med and P1Cin and low to moderate validity for all the other filters if interpreted strictly as Pearson's r correlation coefficients.

When corrected using the maximum value the phi coefficient could take, in order to eliminate the restriction of range of the phi coefficient, all the sensitive (S) strategies exhibit validity coefficients superior to 0.90 (very high validity) in both samples, except for S1Cin, whose coefficient is slightly inferior to 0.90 in Sample 2. The balanced strategies in Medline exhibit a very high validity (phi/phimax coefficient superior to 0.90) in both samples as well, except for SP2Med, whose coefficient is slightly inferior to 0.90 in Sample 1. The balanced strategies in Embase and CINAHL have high validity coefficients ($0.70 \leq \text{phi/phimax} < 0.89$) in both samples. The precise strategies in Embase have moderate validity in both samples ($0.50 \leq \text{phi/phimax} < 0.69$). Most of the precise strategies in Medline and CINAHL have validity varying from low to moderate depending on the sample considered ($0.30 \leq \text{phi/phimax} < 0.69$). This could be due to the fact that these strategies retrieved fewer papers than the number of gold standard papers. Generally speaking, we observed that the more a strategy is sensitive, the more it is valid, even when assessed in a different population (the criterion is the same).

Table 15: Analysis of the predictors: filters developed for Medline, Embase and CINAHL

	Sample 1			Sample 2			χ^2 $\alpha = 5\%$ df = 1	$\frac{P_{Y1}}{P_{Y2}}$
	$(a + b)_1$	P_{Y1}	95% CI	$(a + b)_2$	P_{Y2}	95% CI		
Medline								
<i>S1Med</i>	1003	0.566	[0.543; 0.589]	1823	0.415	[0.401; 0.430]	115.44	1.36
<i>S2Med</i>	282	0.159	[0.143; 0.177]	346	0.079	[0.071; 0.087]	88.96	2.02
<i>S3Med</i>	317	0.179	[0.162; 0.197]	437	0.099	[0.091; 0.109]	73.98	1.80
<i>P1Med</i>	74	0.042	[0.033; 0.052]	131	0.030	[0.025; 0.035]	5.57	1.40
<i>P2Med</i>	111	0.063	[0.052; 0.075]	100	0.023	[0.019; 0.028]	60.64	2.75
<i>P3Med</i>	98	0.055	[0.046; 0.067]	79	0.018	[0.014; 0.022]	62.97	3.07
<i>SP1Med</i>	174	0.098	[0.085; 0.113]	190	0.043	[0.038; 0.050]	68.47	2.27
<i>SP2Med</i>	167	0.094	[0.081; 0.109]	187	0.043	[0.037; 0.049]	62.17	2.21
<i>SP3Med</i>	192	0.108	[0.095; 0.124]	212	0.048	[0.042; 0.055]	74.31	2.24
Embase								
<i>S1Emb</i>	263	0.181	[0.162; 0.202]	292	0.096	[0.086; 0.107]	66.09	1.89
<i>S2Emb</i>	258	0.178	[0.159; 0.198]	289	0.095	[0.085; 0.106]	63.12	1.87
<i>S3Emb</i>	271	0.187	[0.167; 0.208]	336	0.110	[0.100; 0.122]	49.08	1.69
<i>P1Emb</i>	114	0.079	[0.066; 0.094]	70	0.023	[0.018; 0.029]	77.29	3.42
<i>P2Emb</i>	118	0.081	[0.068; 0.096]	70	0.023	[0.018; 0.029]	83.41	3.54
<i>P3Emb</i>	125	0.086	[0.073; 0.102]	68	0.022	[0.018; 0.028]	97.35	3.86
<i>SP1Emb</i>	150	0.103	[0.089; 0.120]	121	0.040	[0.033; 0.047]	70.21	2.60
<i>SP2Emb</i>	158	0.109	[0.094; 0.126]	151	0.050	[0.042; 0.058]	53.94	2.20
<i>SP3Emb</i>	189	0.130	[0.114; 0.148]	131	0.043	[0.036; 0.051]	113.03	3.03
CINAHL								
<i>S1Cin</i>	306	0.309	[0.281; 0.339]	278	0.128	[0.114; 0.143]	148.57	2.42
<i>S2Cin</i>	316	0.319	[0.291; 0.349]	293	0.135	[0.121; 0.150]	149.00	2.37
<i>S3Cin</i>	299	0.302	[0.274; 0.331]	245	0.113	[0.100; 0.127]	171.42	2.68
<i>P1Cin</i>	13	0.013	[0.007; 0.023]	6	0.003	[0.001; 0.006]	12.27	4.76
<i>P2Cin</i>	40	0.040	[0.030; 0.055]	32	0.015	[0.010; 0.021]	20.20	2.75
<i>P3Cin</i>	40	0.040	[0.030; 0.055]	36	0.017	[0.012; 0.023]	16.52	2.44
<i>SP1Cin</i>	97	0.098	[0.081; 0.118]	121	0.056	[0.047; 0.066]	19.02	1.76
<i>SP2Cin</i>	132	0.133	[0.114; 0.156]	151	0.069	[0.059; 0.081]	34.13	1.92
<i>SP3Cin</i>	124	0.125	[0.106; 0.147]	118	0.054	[0.045; 0.065]	48.57	2.31

Table 16: Validity coefficients in the two samples used for the validation study

Search filter	Sample 1				Sample 2			
	ϕ_{XY}	χ^2 ($N\phi_{XY}^2$)	PHI_{MAX}	ϕ_{XY}/PHI_{MAX}	ϕ_{XY}	χ^2 ($N\phi_{XY}^2$)	PHI_{MAX}	ϕ_{XY}/PHI_{MAX}
Medline								
<i>S1Med</i>	0.19	67	0.19	1.00	0.14	89	0.14	1.00
<i>S2Med</i>	0.50	447	0.51	0.99	0.39	673	0.42	0.95
<i>S3Med</i>	0.47	388	0.47	0.99	0.35	520	0.37	0.95
<i>P1Med</i>	0.46	377	0.94	0.49	0.51	1146	0.69	0.74
<i>P2Med</i>	0.57	578	0.86	0.67	0.39	655	0.80	0.49
<i>P3Med</i>	0.53	499	0.92	0.58	0.37	591	0.90	0.41
<i>SP1Med</i>	0.62	677	0.67	0.92	0.54	1279	0.57	0.95
<i>SP2Med</i>	0.60	649	0.69	0.88	0.54	1257	0.58	0.93
<i>SP3Med</i>	0.60	642	0.64	0.95	0.51	1140	0.54	0.95
Embase								
<i>S1Emb</i>	0.51	378	0.51	1.00	0.41	504	0.43	0.94
<i>S2Emb</i>	0.51	375	0.52	0.98	0.41	510	0.44	0.94
<i>S3Emb</i>	0.47	320	0.50	0.94	0.37	413	0.40	0.92
<i>P1Emb</i>	0.52	388	0.82	0.63	0.48	695	0.92	0.52
<i>P2Emb</i>	0.52	389	0.81	0.64	0.48	695	0.92	0.52
<i>P3Emb</i>	0.51	379	0.78	0.65	0.51	765	0.94	0.54
<i>SP1Emb</i>	0.55	435	0.71	0.77	0.56	949	0.70	0.81
<i>SP2Emb</i>	0.54	424	0.69	0.79	0.54	884	0.62	0.88
<i>SP3Emb</i>	0.53	408	0.62	0.85	0.56	949	0.67	0.84
CINAHL								
<i>S1Cin</i>	0.37	133	0.37	1.00	0.41	354	0.41	1.00
<i>S2Cin</i>	0.35	120	0.36	0.97	0.38	304	0.40	0.96
<i>S3Cin</i>	0.35	123	0.37	0.95	0.37	294	0.44	0.85
<i>P1Cin</i>	0.28	77	0.47	0.59	0.28	165	0.34	0.83
<i>P2Cin</i>	0.44	190	0.84	0.52	0.46	441	0.78	0.58
<i>P3Cin</i>	0.39	154	0.84	0.47	0.43	388	0.83	0.52
<i>SP1Cin</i>	0.54	286	0.74	0.72	0.58	711	0.64	0.90
<i>SP2Cin</i>	0.52	269	0.62	0.84	0.56	668	0.57	0.98
<i>SP3Cin</i>	0.51	263	0.65	0.80	0.55	635	0.65	0.84

Discussion

To our knowledge, this study provides the first evidence reporting the concurrent validity of subject bibliographic search filters. Despite the very good validity displayed by most of the filters in their current state, it is our intention to build on that experience to enhance their quality and provide new filters that will be of higher quality.

One of the lessons learned is that when developing optimized bibliographic search filters, special attention should be paid to the gold standard papers selection process. The instrument developed for this selection must be reliable (yielding consistent results in repeated trials) ([Carmines and Woods 2005a](#)) and show very good content validity (adequately covering a specific domain of content) ([Carmines and Woods 2005b](#)).

The systematic reviews software we used to automate the selection process provided kappa coefficients as a measure of the inter-rater reliability ([O'Blenis 2004](#)). We used this feature to monitor the consistency with which our reviewers applied the selection forms to reach greater agreement. However, we did not formally evaluate the reliability of the instruments we developed for the selection of the gold standard papers. Theoretically, with such an evaluation we could have measured not only equivalence in terms of inter-rater agreement, but also stability (whether the same papers would be selected if the process were repeated) and internal consistency (whether the different criteria chosen were related to each other) ([Carmines and Woods 2005a](#)) to ensure that papers meeting our criteria were not excluded inadvertently.

We also did not formally evaluate the content validity of the instruments developed for the gold standard papers selection. For example, we could have asked the reviewers to provide reasons for exclusion, to make sure they did not agree to exclude papers that could have been included as gold standard papers. The list of keywords provided as indicators of gold standard papers could have been validated by content experts in the field. Therefore, it is possible that papers that could have been worth selecting as gold standard papers were excluded.

Another lesson learned concerns the samples of papers corresponding to the two sets of gold standard papers (Sample 1 and Sample 2). The fact that the gold standard papers were selected according to the same set of criteria does not mean that any method used to allocate them into two groups as well as to create the corresponding samples of papers will produce comparable groups. Our initial goal was to demonstrate the stability of the relationship between the predictors and the criterion by using a sample different from the one used to build the filters. However, when the goal is to conduct a double validation (replication or cross validation study), the sampling method used should allow for two independent groups to come from the same population in order to compare the validity coefficients obtained while controlling for random errors ([Harter 1986](#); [Pettersen and Ebrary 2000](#)). In our case, the two samples were significantly different, in terms of the proportions both of gold standard papers and of papers retrieved by the strategies. Therefore, we showed that the sensitive strategies and the balanced strategies have good criterion-related validity even in different populations, but it was difficult to compare the coefficients obtained for most of the precise strategies, as both systematic differences and random errors were included. We also noticed that the proportions of papers retrieved by the precise filters, whose validity coefficients were the most difficult to compare over the two samples, were most often inferior to the proportion of gold standard papers in the samples. This could indicate that, when building precise strategies, attention should be paid not only to maximizing precision but also to the level of sensitivity obtained; in order to make sure these strategies will exhibit good validity as well. This difficulty in comparing coefficients could also be linked to the choice of an appropriate sample size for the gold standards and their corresponding samples of papers.

Conclusion

The most sensitive and balanced strategies can be validly interpreted as a measure of papers being retrieved dealing with patient safety in health care organizations. They appear to remain valid instruments even when the populations of articles considered are significantly different in terms of the proportion of gold standard papers or the proportion of papers retrieved by the filters. Based on these results, the authors encourage more

research on the appropriate sampling method for search filters construction and validation, in order to support double validation as well as to support generalization studies. Instruments used to build the set of gold standard papers should be assessed for reliability and content validity as part of the criterion-related validity argumentation. The more sensitive strategies should be considered first when trying to apply the filters outside the experimental context of their construction.

Competing interests

The authors declare that they have no competing interests.

Contributors

Affaud Anaïs Tanon, François Champagne, and André-Pierre Contandriopoulos all contributed to the conception and design of the study. Affaud Anaïs Tanon was responsible for the acquisition, analysis, and interpretation of the data, and drafted the article. All authors were involved in critically reviewing the article and gave final approval of the version submitted for publication.

Acknowledgments

This research was funded by the Canadian Institutes of Health Research (CIHR).

Scholarships were provided by: the AnÉIS program (Analyse et Évaluation des Interventions en Santé / Analysis and Evaluation of Health Interventions); SSHRC (Social Sciences and Humanities Research Council of Canada); GIRU (Groupe Interuniversitaire de Recherche sur les Urgences / Interuniversity Research Group on Emergency Services); the Government of Côte d'Ivoire; the University of Montreal; and a research project funded by CIHR (Canadian Institutes of Health Research).

The project was conducted with the support of the Takemi Program in International Health at Harvard School of Public Health. We would like to thank

Professors Michelle Mello and Ashish Jha, advisers to Affaud Anaïs Tanon at Harvard University. We also thank Su Golder at the University of York and Professor Kaveh Shojania at the University of Toronto for their comments on previous drafts.

Chapitre 7: L'utilité des filtres de recherche proposés pour Medline, Embase et CINAHL.

Article 4: Sécurité des Patients et revues systématiques : utilité des filtres de recherche optimisés construits pour Medline, Embase and CINAHL.

Anaïs Tanon, André-Pierre Contandriopoulos, François Champagne

Introduction

Il est clairement établi dans la littérature que les filtres de recherche bibliographique optimisés ont tendance à être moins performants en situation réelle d'utilisation que dans les conditions expérimentales dans lesquelles ils ont été développés ([Jenkins 2004](#)) . Les filtres de recherche optimisés sont des combinaisons logiques de mots clés et/ou de termes d'indexation construites pour repérer des études effectuées selon une certaine méthodologie ou la littérature pertinente sur un sujet donné dans une base de données particulière. Afin que les utilisateurs potentiels des filtres de recherche puissent apprécier l'utilité des filtres de recherche bibliographique optimisés qui leur sont proposés, il est souvent recommandé aux concepteurs de filtres de tester la performance de leurs filtres dans un ensemble d'articles (un gold standard) élaboré selon des règles différentes de celles qui ont prévalu lors de la construction des filtres ([Glanville, Bayliss et al. 2008](#); [Bak, Mierzewski-Urban et al. 2009](#)). Cette façon de faire ne donne pas toujours des résultats convergents quant à l'utilité des filtres de recherche car la performance varie énormément en fonction du groupe d'articles utilisé pour l'exercice de validation ([Leeflang, Scholten et al. 2006](#); [Ritchie, Glanville et al. 2007](#); [Kastner, Wilczynski et al. 2009](#); [Whiting, Westwood et al. 2011](#)).

Une méthode plus appropriée pour mesurer la validité des filtres de recherche, c'est-à-dire la mesure dans laquelle l'interprétation que l'on fait des résultats des filtres de recherche pour un usage particulier est soutenue par les évidences et par la théorie ([American Educational Research Association, American Psychological Association et al. 1999](#)), consisterait par exemple à mesurer la validité de critère, comme cela est pratique courante pour les instruments de mesure ([Nunnally and Bernstein 1994](#); [Anastasi and Urbina 1997](#); [Carmines and Woods 2005b](#)). L'exercice fournit des coefficients de validité qui mesurent le degré de certitude avec lequel on peut affirmer qu'un article sélectionné à l'aide d'un filtre de recherche l'aurait également été si on avait pu mesurer directement le critère d'intérêt ([Carmines and Woods 2005b](#)). La validation de critère se penche également sur la validité des décisions de sélection prises à l'aide du filtre de recherche, c'est-à-dire sur la mesure dans laquelle les classifications obtenues à l'aide du filtre de recherche correspondent aux classifications qui auraient été obtenues si on avait mesuré directement le critère d'intérêt ([Anastasi and Urbina 1997](#)), ce qui permet de fournir un éclairage sur l'utilité potentielle d'un filtre de recherche.

La non convergence des résultats obtenus par la méthode qui consiste à simplement mesurer la performance des filtres dans un autre groupe d'articles peut s'expliquer par le fait que la validité d'un instrument de mesure est situationnelle ([American Educational Research Association, American Psychological Association et al. 1999](#)). En effet la validité n'est pas une propriété intrinsèque du filtre de recherche mais est plutôt liée à l'interprétation que l'on fait des résultats obtenus grâce au filtre de recherche dans un contexte particulier ([Furr and Bacharach 2008](#)). Cependant, cette méthode peut s'avérer pertinente si l'on cherche à savoir si l'outil pourrait être utile dans une situation où l'on a par ailleurs de fortes raisons de penser que les filtres devraient être performants.

Nous avons proposé, dans un précédent article, des filtres de recherche élaborés entre 2007 et 2008 pour repérer des articles portant sur la sécurité des patients ([Tanon, Champagne et al. 2010](#)). Ces filtres ont été élaborés pour repérer de façon plus spécifique les articles portant sur la sécurité des patients dans les organisations de santé qui ont été définis de façon explicite comme pertinents pour le champ de la sécurité des patients par

leurs auteurs dans un souci de repérer des articles qui permettraient de mieux comprendre le concept de Sécurité des Patients.

Des articles ont été choisis pour former deux échantillons de convenance afin d'avoir deux gold standards qui ont été utilisés pour construire les filtres de recherche d'une part et pour évaluer leur performance d'autre part ([Tanon, Champagne et al. 2010](#)).

Mais le champ de la sécurité des patients étant encore en plein processus de maturation ([Leape 2008](#); [Emanuel, Berwick et al. 2008](#)), nous étions intéressés à savoir si les filtres élaborés pourraient s'avérer utiles dans un contexte où le contenu du terme « Patient Safety » s'avère plus large que celui considéré lors de la construction des filtres de recherche. Pour ce faire entre 2009 et 2010, nous avons élaboré un nouveau gold standard composé d'articles portant sur la sécurité des patients mais choisis selon des critères différents afin de déterminer si parmi les filtres élaborés, certains pourraient s'avérer plus utiles dans un contexte plus large.

Ce troisième échantillon d'articles a été construit à partir des suggestions de lecture de la National Patient Safety Foundation (ci-après NPSF), une organisation américaine créée en 1997 pour améliorer la sécurité des patients. Deux fois par mois depuis 1997, la NPSF-US publie une liste annotée de ressources pertinentes sur la sécurité des patients, nommée « Current Awareness » ([National Patient Safety Foundation](#)) ([National Patient Safety Foundation resource page](#)). Cette publication n'est accessible désormais qu'aux membres de l'American Society of Professionals in Patient Safety. Nous l'avons choisi pour construire un nouveau gold standard pour trois raisons principales. Tout d'abord, la possibilité de couvrir toutes les années depuis 1999, date à partir de laquelle la littérature sur la Sécurité des patients a pratiquement explosé, tant en quantité, qu'en étendue ([Stelfox, Palmisani et al. 2006](#)). Ensuite, les suggestions renvoient non seulement à des articles scientifiques, mais aussi à d'autres types de ressources, ce qui donne une idée plus large du champ de la sécurité des patients. Enfin, cette organisation a été créée par les pionniers du champ de la sécurité des patients ([Spiegel and Kavalier 1997](#); [Leape 2008](#)), ce qui fournit un jugement valide (quoiqu'implicite) sur la pertinence des ressources collectées pour le champ de la sécurité des patients.

Pour construire le nouveau gold standard ci-après nommé NPSF_GS, la liste de ressources publiées entre Janvier 1999 et Décembre 2008 a été épurée pour éliminer les doublons et reclasser les articles selon l'année de publication ou de parution. Par la suite, nous avons sélectionné les éléments de cette liste, indexés dans au moins l'une des trois bases de données Medline, Embase et CINAHL. La recherche s'est faite en utilisant différents champs d'indexation : titre de l'article, l'auteur, le titre des journaux, les issn, etc.

L'objectif de cet article est double. Dans un premier temps, il s'agit d'évaluer l'utilité des filtres de recherche à effectuer les tâches de repérage pour lesquelles ils ont été élaborés à l'aide des deux premiers échantillons qui ont servi à leur construction et à l'évaluation de leur performance. Dans un deuxième temps, il s'agit de tester la performance des filtres de recherche élaborés à l'aide du troisième échantillon d'articles, le gold standard NPSF_GS afin de repérer ceux des filtres proposés qui pourraient être utiles dans un contexte plus large.

Méthode

Descriptions des gold standards utilisés :

Le premier gold standard (Échantillon 1) est composé d'articles publiés dans JAMA (2001), Quality & Safety in Health Care (2004, 2005 and 2006) et Nursing Management (2005). Le deuxième gold standard (Échantillon 2) est composé d'articles publiés dans BMJ (2000), the New England Journal of Medicine (2003), Quality & Safety in Health Care (2002 and 2003) and Nursing Management (2006). La sélection des articles composant le gold standard s'est faite en deux étapes selon les critères suivants :

- Les articles doivent mentionner de façon explicite des termes qui renvoient à la sécurité des patients tels que patient safety, medical errors, medication error(s), diagnostic error(s), adverse event(s), adverse effect(s), adverse outcome(s), undesirable event(s), undesirable effect(s), undesirable outcome(s), medical injury(ies), etc...et les termes qui leur sont proches. Ce choix était justifié par le

souci de se concentrer dans un premier temps sur les textes dont les auteurs s’inscrivaient de façon délibérée dans le champ de la sécurité des patients par l’utilisation de termes consacrés afin de procéder par la suite à une analyse de l’évolution du champ;

- Les articles parlent explicitement de la sécurité des patients (même s’ils peuvent aborder la sécurité d’autres groupes tels que les prestataires de soins);
- Les articles discutent de la sécurité des patients dans le contexte des organisations de santé.

Table 17 : Les deux échantillons pour l’analyse de l’utilité.

Base de donnée	Échantillon 1		Échantillon 2	
	Taille de l’ensemble d’articles	Taille du gold standard	Taille de l’ensemble d’articles	Taille du gold standard
Medline	1773	83	4392	64
Embase	1452	79	2146	60
CINAHL	990	56	2175	52

Le troisième gold standard (Échantillon 3) est composé de 3540 références publiées entre 1999 et 2008 dans plus de 500 journaux (Tableau 18). Les critères qui ont guidé la sélection des articles composant ce gold standard sont très différents de ceux qui ont servi à sélectionner les articles des gold standard 1 et 2. En effet les termes classiques tel que “patient safety” “medical errors” “medication errors” “adverse or undesirable events, effects or outcomes”, “medical injuries,” etc ont été utilisés dans un premier temps. Par la suite, le processus de sélection a évolué. Les concepts relatifs à la science des systèmes, aux systèmes de sécurité, aux facteurs humains, aux histoires de patients, aux expériences des cliniciens et aux façons dont tous ces éléments pourraient être convertis en ressources commerciales ont été repérés dans les actualités, les bases de données en sciences médicales et en sciences fondamentales. À partir de 2005, le concept plus large de la qualité a commencé à être incorporé comme critère de sélection (Lorri Zipperer, correspondances personnelles Avril 2011).

Tableau 18 : Les entités listées dans la publication « Current Awareness » de la US National Patient Safety Foundation entre 1999 et 2008.

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	TOTAL
Records indexed in at least one of the databases.	146	370	470	480	453	389	333	289	292	318	3540
Records indexed in MEDLINE	129	323	415	439	430	373	314	273	277	281	3254
Records indexed in EMBASE	102	259	271	317	307	273	207	179	171	208	2294
Records indexed in CINAHL	50	115	192	235	246	190	196	232	239	200	1895
Records indexed in none of the three databases		129	295	254	155	120	141	109	65	92	94
Audio Media	3	2	5	2	0	0	1	0	2	1	16
Newspaper	50	147	93	32	4	1	1	2	2	9	341
Magazines articles	25	47	39	30	25	10	37	25	27	10	275
Books	3	5	4	3	5	11	1	1	1	1	35
Scientific Papers	24	40	42	26	12	18	26	13	23	32	256
Grey literature	24	54	71	62	74	101	43	24	37	41	531
Total number of records	275	665	724	635	573	530	442	354	384	412	4994

En comparaison du premier gold standard qui a servi à construire les filtres de recherche, le troisième gold standard contient plus d'articles (3540 versus 85⁵), issus d'un plus grand nombre de journaux (plus de 500, versus 3), publiés sur une plus grande période de temps (1999-2008, versus 2001, 2004, 2005 et 2006). Le gold standard NPSF a une plus grande diversité en termes de types d'articles, en termes de pays d'origine des premiers auteurs, et en termes d'institutions représentées (Annexe 8).

Le gold standard NPSF n'est pas un ensemble exhaustif de tous les articles pertinents dans le champ de la Sécurité des Patients, il est composé d'articles intéressants que le Centre des Ressources en Information de la National Patient Safety Foundation a jugé bon de porter à l'attention de la Communauté des personnes intéressées par la Sécurité des Patients. En comparant les deux collections d'articles, nous avons pu constater que les deux gold standard ont en commun seulement 21 articles. Le gold standard NPSF constitue donc un choix intéressant pour tester la performance des filtres en situation non expérimentale car il comprend, un grand nombre des facteurs reconnus pour limiter la capacité des filtres de recherche à reproduire les mêmes résultats que ceux

⁵ Dont 83 sont indexés dans Medline.

obtenus lors de leur construction ([Bachmann, Coray et al. 2002](#); [Jenkins 2004](#); [Leeflang, Scholten et al. 2006](#); [Wong, Wilczynski et al. 2006](#)): plus d'années de publication couvertes, une plus grande étendue de journaux, de types de publication et de disciplines couvertes et peu d'articles en commun.

Mesurer la validité des sélections faites à l'aide des filtres de recherche.

La mesure de la validité des sélections faites à l'aide des filtres de recherche sera conduite avec les échantillons 1 et 2. La matrice de classification présentée dans le Tableau 19 sert à évaluer la performance et la validité des filtres de recherche et décrit les décisions qu'un filtre de recherche peut permettre de prendre lorsqu'utilisé comme instrument de sélection dans un ensemble d'articles :

- Les vrais positifs (**a**) : les articles qui sont repérés par le filtre de recherche et qui appartiennent au gold standard;
- Les faux positifs (**b**) : les articles qui sont repérés par le filtre de recherche et qui n'appartiennent pas au gold standard;
- Les faux négatifs (**c**) : les articles qui ne sont pas repérés par le filtre de recherche mais qui auraient dû l'être car ils appartiennent au gold standard;
- Les vrais négatifs (**d**) : les articles qui ne sont pas repérés par le filtre de recherche à juste titre car ils n'appartiennent pas au gold standard;
- Les bonnes décisions (**a+d**) : la somme des vrais positifs et des vrais négatifs;
- Les mauvaises décisions (**b + c**) : la somme des faux positifs et des faux négatifs.

Un filtre de recherche bibliographique sera d'autant plus utile qu'il permettra de prendre de bonnes décisions. Il existe plusieurs indices statistiques permettant de mesurer l'efficacité de la prédiction des instruments de mesure tels que le kappa, l'index d'efficacité prédictive, la statistique RIOC ([Farrington and Loeber 1989](#); [Smith 1996](#)). Notre choix s'est porté sur la statistique RIOC (Relative Improvement Over Chance) ([Copas and Loeber 1990](#)) car elle permet non seulement de corriger pour la chance mais aussi pour le fait que dans les tableaux de contingence, une grande différence entre les valeurs moyennes des deux variables considérées limite sérieusement les possibilités de

classification. Cette situation est celle qui prévaut dans les deux échantillons qui ont servi à la construction et à l'évaluation de la performance des filtres et que nous utilisons également pour démontrer l'utilité des filtres de recherche. Dans chacun des échantillons, il existe une grande différence entre la proportion des articles définis comme faisant partie du gold standard et la proportion des articles repérés par les filtres (Tableau 20 et Tableau 21) .

La statistique du RIOC varie entre 0 et 1 et présente le pourcentage de bonnes décisions prises en utilisant un instrument de mesure qui excède le pourcentage des bonnes décisions qui auraient pu être prises en sélectionnant les articles de façon aléatoire ([Copas and Loeber 1990](#); [Cairney and Streiner 2011](#)). Le Tableau 19 présente les formules utilisées pour calculer le RIOC, son intervalle de confiance et son niveau de signification.

Tableau 19 : Formules pour calculer le RIOC (Relative Improvement Over Chance).

Filtre de recherche	Échantillon d'articles		Total
	Articles repérés	Articles non repérés	
	a	b	R ₁ = (a + b)
	c	d	R ₂ = (c + d)
Total	C ₁ = (a + c)	C ₂ = (b + d)	N = (a + b + c + d)

Statistique:	Formule:
RIOC (Relative Improvement Over Chance)	$RIOC = \frac{(a+d) - \left(\frac{(a+c)(a+b)}{N} + \frac{(b+d)(c+d)}{N} \right)}{\max(a+d) - \left(\frac{(a+c)(a+b)}{N} + \frac{(b+d)(c+d)}{N} \right)} * 100$
Intervalle de confiance pour le RIOC	<p>$RIOC \pm S_{RIOC}$ où:</p> $S_{RIOC} = \sqrt{S_{RIOC}^2}$ <p>et $S_{RIOC}^2 = \frac{Nc(NC_1R_2 + c(NR_1 + R_1C_1 - 2NC_1 - N^2) + 2Nc^2)}{R_2^3 C_1^3}$</p> <p>$R_1 = (a + b); R_2 = (c + d); C_1 = (a + c); C_2 = (b + d)$</p>
(RIOC S[*]_{RIOC})	$(RIOC S_{RIOC}^*) = \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ <p>Pour tester la différence entre le RIOC et la valeur de chance 0, la quantité $(RIOC S_{RIOC}^*)$ est testée comme une χ^2 à degré de liberté.</p>
Statistique de comparaison des RIOCs	$\frac{RIOC_1 - RIOC_2}{\sqrt{RIOC_1^2 - RIOC_2^2}}$ <p>La différence entre les deux RIOCs est significative à un degré de signification de 95% si la statistique est supérieure à 2 en valeur absolue</p>

Tableau 20 : Statistiques de base, taux de base et taux de sélection pour les filtres de recherche dans l'échantillon 1 qui a servi à la construction des filtres de recherche.

	<i>Vrais positifs (a)</i>	<i>Faux positifs (b)</i> <i>n</i>	<i>Faux positifs (b)</i> <i>%</i>	<i>Faux négatifs (c)</i> <i>n</i>	<i>Faux négatifs (c)</i> <i>%</i>	<i>Vrais négatifs (d)</i>	<i>N</i>	<i>Taux de base</i>	<i>Taux de sélection</i>
Medline									
<i>S1Med</i>	83	920	51.89%	0	0.00%	770	1773	4.68%	56.57%
<i>S2Med</i>	82	200	11.28%	1	0.06%	1490	1773	4.68%	15.91%
<i>S3Med</i>	82	235	13.25%	1	0.06%	1455	1773	4.68%	17.88%
<i>P1Med</i>	38	36	2.03%	45	2.54%	1654	1773	4.68%	4.17%
<i>P2Med</i>	57	54	3.05%	26	1.47%	1636	1773	4.68%	6.26%
<i>P3Med</i>	50	48	2.71%	33	1.86%	1642	1773	4.68%	5.53%
<i>SP1Med</i>	77	97	5.47%	6	0.34%	1593	1773	4.68%	9.81%
<i>SP2Med</i>	74	93	5.25%	9	0.51%	1597	1773	4.68%	9.42%
<i>SP3Med</i>	79	113	6.37%	4	0.23%	1577	1773	4.68%	10.83%
Embase									
<i>S1Emb</i>	79	184	12.67%	0	0.00%	1189	1452	5.44%	18.11%
<i>S2Emb</i>	78	180	12.40%	1	0.07%	1193	1452	5.44%	17.77%
<i>S3Emb</i>	75	196	13.50%	4	0.28%	1177	1452	5.44%	18.66%
<i>P1Emb</i>	52	62	4.27%	27	1.86%	1311	1452	5.44%	7.85%
<i>P2Emb</i>	53	65	4.48%	26	1.79%	1308	1452	5.44%	8.13%
<i>P3Emb</i>	54	71	4.89%	25	1.72%	1302	1452	5.44%	8.61%
<i>SP1Emb</i>	63	87	5.99%	16	1.10%	1286	1452	5.44%	10.33%
<i>SP2Emb</i>	64	94	6.47%	15	1.03%	1279	1452	5.44%	10.88%
<i>SP3Emb</i>	69	120	8.26%	10	0.69%	1253	1452	5.44%	13.02%
CINAHL									
<i>S1Cin</i>	56	250	25.25%	0	0.00%	684	990	5.66%	30.91%
<i>S2Cin</i>	55	261	26.36%	1	0.10%	673	990	5.66%	31.92%
<i>S3Cin</i>	54	245	24.75%	2	0.20%	689	990	5.66%	30.20%
<i>P1Cin</i>	8	5	0.51%	48	4.85%	929	990	5.66%	1.31%
<i>P2Cin</i>	22	18	1.82%	34	3.43%	916	990	5.66%	4.04%
<i>P3Cin</i>	20	20	2.02%	36	3.64%	914	990	5.66%	4.04%
<i>SP1Cin</i>	42	55	5.56%	14	1.41%	879	990	5.66%	9.80%
<i>SP2Cin</i>	48	84	8.48%	8	0.81%	850	990	5.66%	13.33%
<i>SP3Cin</i>	46	78	7.88%	10	1.01%	856	990	5.66%	12.53%

Tableau 21 : Statistiques de base, taux de base et taux de sélection pour les filtres de recherche dans l'échantillon 2 qui a servi à l'évaluation de la performance des filtres de recherche.

	<i>Vrais positifs</i>	<i>Faux positifs</i>		<i>Faux négatifs</i>		<i>Vrais négatifs</i>	<i>Total</i>	<i>Taux de base</i>	<i>Taux de sélection</i>	
		<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>					
Medline										
<i>S1Med</i>	64	1759	40.05%	0	0.00%	2569	4392	1.46%	41.51%	
<i>S2Med</i>	61	285	6.49%	3	0.07%	4043	4392	1.46%	7.88%	
<i>S3Med</i>	61	376	8.56%	3	0.07%	3952	4392	1.46%	9.95%	
<i>P1Med</i>	48	83	1.89%	16	0.36%	4245	4392	1.46%	2.98%	
<i>P2Med</i>	32	68	1.55%	32	0.73%	4260	4392	1.46%	2.28%	
<i>P3Med</i>	27	52	1.18%	37	0.84%	4276	4392	1.46%	1.80%	
<i>SP1Med</i>	61	129	2.94%	3	0.07%	4199	4392	1.46%	4.33%	
<i>SP2Med</i>	60	127	2.89%	4	0.09%	4201	4392	1.46%	4.26%	
<i>SP3Med</i>	61	151	3.44%	3	0.07%	4177	4392	1.46%	4.83%	
Embase										
<i>S1Emb</i>	57	235	7.72%	3	0.10%	2751	3046	1.97%	9.59%	
<i>S2Emb</i>	57	232	7.62%	3	0.10%	2754	3046	1.97%	9.49%	
<i>S3Emb</i>	56	280	9.19%	4	0.13%	2706	3046	1.97%	11.03%	
<i>P1Emb</i>	32	38	1.25%	28	0.92%	2948	3046	1.97%	2.30%	
<i>P2Emb</i>	32	38	1.25%	28	0.92%	2948	3046	1.97%	2.30%	
<i>P3Emb</i>	33	35	1.15%	27	0.89%	2951	3046	1.97%	2.23%	
<i>SP1Emb</i>	49	72	2.36%	11	0.36%	2914	3046	1.97%	3.97%	
<i>SP2Emb</i>	53	98	3.22%	7	0.23%	2888	3046	1.97%	4.96%	
<i>SP3Emb</i>	51	80	2.63%	9	0.30%	2906	3046	1.97%	4.30%	
CINAHL										
<i>S1Cin</i>	52	226	10.39%	0	0.00%	1897	2175	2.39%	12.78%	
<i>S2Cin</i>	50	243	11.17%	2	0.09%	1880	2175	2.39%	13.47%	
<i>S3Cin</i>	45	200	9.20%	7	0.32%	1923	2175	2.39%	11.26%	
<i>P1Cin</i>	5	1	0.05%	47	2.16%	2122	2175	2.39%	0.28%	
<i>P2Cin</i>	19	13	0.60%	33	1.52%	2110	2175	2.39%	1.47%	
<i>P3Cin</i>	19	17	0.78%	33	1.52%	2106	2175	2.39%	1.66%	
<i>SP1Cin</i>	47	74	3.40%	5	0.23%	2049	2175	2.39%	5.56%	
<i>SP2Cin</i>	51	100	4.60%	1	0.05%	2023	2175	2.39%	6.94%	
<i>SP3Cin</i>	44	74	3.40%	8	0.37%	2049	2175	2.39%	5.43%	

Tester les filtres de recherche à l'aide du gold standard NPSF.

Lors de la construction et de l'évaluation de la performance des filtres, cinq indicateurs ont été mesurés ([Tanon, Champagne et al. 2010](#)) : la sensibilité, la spécificité, la précision, le produit de la sensibilité et de la précision, et le nombre d'articles à lire qui est l'inverse de la précision⁶. Quatre pièces d'information ont permis leur calcul : le nombre d'articles pertinents extraits par les filtres de recherche, le nombre d'articles pertinents non extraits par les filtres de recherche, le nombre d'articles non pertinents extraits par les filtres de recherche, le nombre d'articles non pertinents et non extraits par les filtres de recherche.

De façon idéale, les mêmes indicateurs auraient dû servir pour l'évaluation de la performance à l'aide du gold standard NPSF. Mais l'ensemble des documents non pertinents n'étant pas délimité, il n'était pas possible de calculer la précision et la spécificité. Il manquait donc des informations clés dans le tableau de contingence correspondant au gold standard NPSF comme le montre le Tableau 22 qui la matrice de classification pour le calcul de la performance des filtres de recherche à l'aide du gold standard NPSF sans considération d'une base de données particulière..

En effet les documents du NPSF-GS sont extraits de plus de 500 journaux mais ne constituent pas un ensemble exhaustif des publications pertinentes à l'intérieur de cet univers. Par conséquent, la taille totale N des articles contenus dans cet ensemble de journaux est inconnue. De la même façon pour chaque filtre de recherche testé, il n'est pas possible d'avoir le nombre b des articles publiés dans les journaux source et repérés par le filtre de recherche qui ne font pas partie du gold standard NPSF, le nombre d articles publiés dans les journaux source et qui ne sont pas repérés par le filtre de recherche à juste titre car ils ne font pas partie du gold standard NPSF. Cela entraîne que les quantités $(a + b)$, $(c + d)$ et $(b + d)$ ne pouvaient être calculés et si on se réfère aux formules nécessaires au calcul des indicateurs sélectionnés pour mesurer la performance

⁶ **Sensibilité** : le nombre de documents pertinents extraits par les filtres de recherche / le nombre de documents pertinents * 100 ; **Spécificité** : le nombre de documents non pertinents et non extraits par les filtres de recherche / le nombre de documents non pertinents * 100 ; **Précision** : le nombre de documents pertinents extraits par les filtres de recherche / le nombre total de documents extraits par les filtres.

des filtres de recherche présentées dans le Tableau 23 , seule la sensibilité pouvait être calculée à partir des données.

Tableau 22: Matrice de classification pour l'évaluation de la performance des filtres de recherche à l'aide du gold standard NPSF.

		Échantillon d'articles		Total
Filtre de recherche	Articles repérés	Pertinents (gold standard) a= le nombre d'articles du gold standard NPSF repérés par le filtre de recherche X	Non Pertinents b = ?	(a +b) = ?
	Articles non repérés	c = le nombre d'articles du gold standard NPSF non repérés par le filtre de recherche X	d =?	(c+d) = ?
Total		(a+c) = 3540	(b+d) = ?	N = ?

Tableau 23: Indicateurs pour évaluer la performance des filtres de recherche.

Sensibilité	$= a/(a + c)$
Précision	$= a/(a + b)$
Spécificité	$= d/(b + d)$
Nombre d'articles à lire	$= (a + b)/a$
Sensibilité *Précision	$= a((a+c)) * a((a+b)) = a^2/(a + c)(a + b)$

Par contre, nous avons utilisé une autre méthode de calcul pour trouver une mesure qui servirait de proxy pour la précision que nous appelons $Précision_{Est}$. La précision a été estimée à partir d'une méthode à posteriori (post hoc precision) ([Boluyt, Tjosvold et al. 2008](#); [Damarell, Tieman et al. 2011](#)). Les filtres de recherche ont été exécutés dans les 3 bases de données Medline, Embase et CINAHL. Dans chacune des bases de données, nous avons appliqué des limites temporelles pour regrouper les articles repérés par chacun des filtres par année pour la période 1999 – 2008, ce qui nous a permis d'obtenir le nombre d'articles repérés par les filtres de recherche dans leur base de données respective par année. Chaque base de données attribue aux articles repérés un

numéro d'apparition à l'écran. Nous avons donc utilisé un générateur de nombres aléatoires ([Randomness and Integrity Services Limited](#)) pour sélectionner de façon aléatoires 10 articles par année pour chacun des filtres de recherche, ce qui donne un total de 2700 articles (100 articles pour chacun des 27 filtres de recherche) (Annexe 9). Les articles étaient jugés pertinents si :

- Ils appartenaient au gold standard NPSF ;
- Sans appartenir au gold standard NPSF :
 - Ils contenaient dans le titre ou le résumé les termes couverts par les articles du gold standard NPSF tels que «patient safety», «medical errors», «adverse event», «medication errors», «diagnostic errors», «safety and efficacy» of therapies, procedures, treatment, medical techniques, etc., «iatrogenic infections», «nosocomial infections», «adverse effects» of treatment or procedures, etc., «treatment outcomes», «drug safety profile», «complications», «adverse drug reactions», «drug interaction», «Pressure ulcers», «Record keeping», «Root cause analysis» «safety culture »;
 - A la lecture l'article portait sur des thèmes abordés dans le gold standard NPSF tels que: patient role in disease management, workplace safety, occupational safety, patient aggression and violence decision making among therapists, occupational safety, drug safety, health care worker environment, health worker fatigue, violence in the health care setting, language barriers, health services accreditation, environmental safety, patient suicide, team work, etc.
- Les articles difficiles à classer ont été examinés par un deuxième réviseur afin d'arriver à une décision définitive.

À partir de la sensibilité et de la précision estimée ($Précision_{Est}$), nous avons calculé deux proxys pour le produit de la sensibilité par la précision ($Sensibilité * Précision_{Est}$) et pour le nombre d'articles à lire ($Nombre\ d'articles\ à\ lire_{Est} = 1/Précision_{Est}$) qui sont donc eux aussi des statistiques estimées.

Résultats

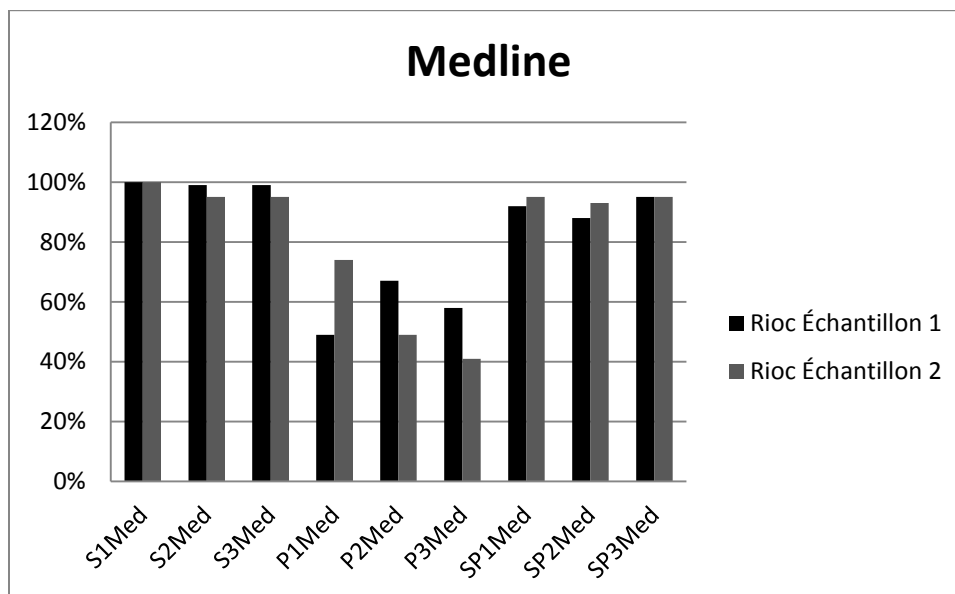
L'analyse de l'utilité des décisions prises à l'aide des filtres de recherche

Le Tableau 24 donne de façon plus détaillée les résultats des statistiques RIOC pour chacun des filtres de recherche dans les échantillons 1 et 2 qui ont servi à la construction et à l'évaluation de la performance des filtres de recherche.

Medline.

Les stratégies les plus sensibles de Medline (S1Med, S2Med et S3Med) de même que les stratégies les plus équilibrées de Medline (SP1Med, SP2Med, SP3Med) améliorent la qualité de la sélection d'articles portant sur la sécurité des patients dans les organisations de santé et inscrits dans le champ de la sécurité des patients de façon explicite par leurs auteurs, de plus de 80% dans les deux échantillons. Il n'y a pas de différences significatives entre les valeurs du RIOC, dans les deux échantillons. Les stratégies les plus précises améliorent elles aussi la qualité de la prise de décision quoique qu'avec moins d'ampleur que les deux autres types de filtres de recherche (les RIOCs varient entre 41% et 74%). La capacité des stratégies les plus précises de Medline à améliorer la prise de décision est fortement liée à l'échantillon d'articles dans lequel les filtres sont testés (Statistique du test de comparaison des RIOC est supérieure à 2 en valeur absolue pour les 3 filtres, ce qui indique que la différence de valeur entre les échantillons est significative) (Figure 6 et Tableau 24).

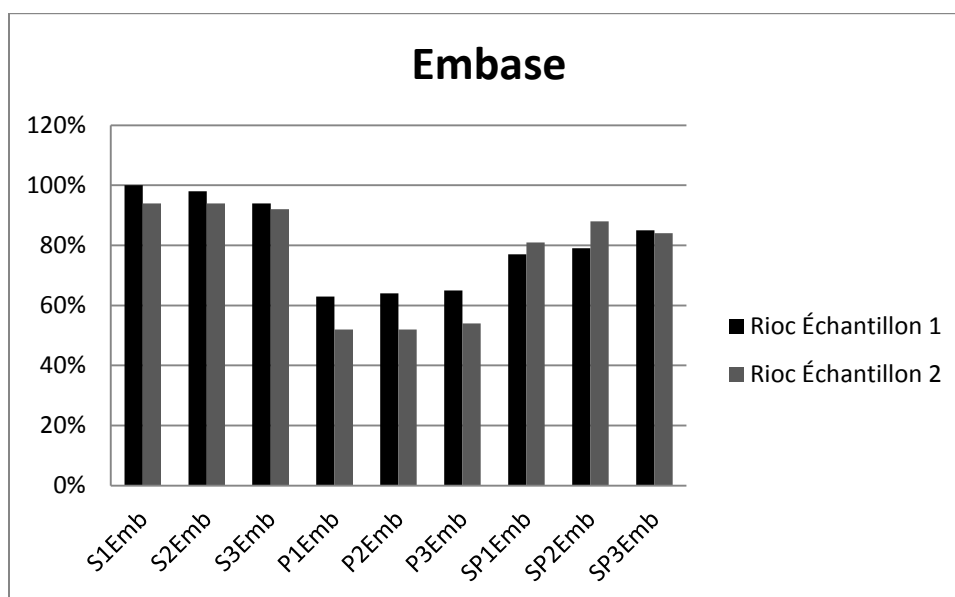
Figure 6 : Amélioration de la prise de décision obtenue pour les filtres de Medline.



Embase.

Dans la base de données Embase, les stratégies les plus sensibles (S1Emb, S2Emb, S3Emb) contribuent à plus de 90% à améliorer la qualité de sélection d'articles répondant aux critères de l'étude dans les deux échantillons. Les stratégies les plus équilibrées (SP1Emb, SP2Emb, SP3Emb) améliorent la qualité de sélection des articles d'environ 80% en moyenne dans les deux échantillons. Les stratégies les plus précises améliorent la qualité de la prise de décision relativement aux articles répondant à nos critères de plus de 50%. Pour toutes les stratégies de Embase, il n'y a pas de différences significatives entre les deux échantillons. (Figure 7 et Tableau 24)

Figure 7 : Amélioration de la prise de décision obtenue pour les filtres d'Embase.



CINAHL.

Les stratégies les plus sensibles de CINAHL (S1Cin, S2Cin et S3Cin) améliorent la qualité de la sélection d'articles répondant à nos critères en moyenne de 95% dans les deux échantillons. Il n'y a pas de différences significatives entre les deux échantillons. La stratégie la plus précise de CINAHL (P1Cin) améliore significativement plus la prise de décision dans l'échantillon 2 (RIOCI = 83%) que dans l'échantillon 1 (RIOCI = 59%), (la statistique du test de comparaison des RIOCI est supérieure à 2 en valeur absolue). Les deux autres stratégies les plus précises améliorent la qualité de la sélection d'articles de plus de 50% en moyenne, mais les différences entre les deux échantillons ne sont pas significatives. Les deux stratégies les plus équilibrées de CINAHL (SP1Cin et SP2Cin) améliorent significativement plus la prise de décision dans l'échantillon 2 (RIOCI = 98% et 90% respectivement) que dans l'échantillon 1 (RIOCI = 72% et 84% respectivement). La statistique du test de comparaison des RIOCI est supérieure à 2 en valeur absolue. La stratégie SP3Cin améliore la qualité de la prise de décision d'environ 80% dans les deux échantillons. Il n'y a pas de différence significative entre les deux échantillons. (Figure 8 et Tableau 24).

Figure 8 : Amélioration de la prise de décision obtenue pour les filtres de CINAHL.

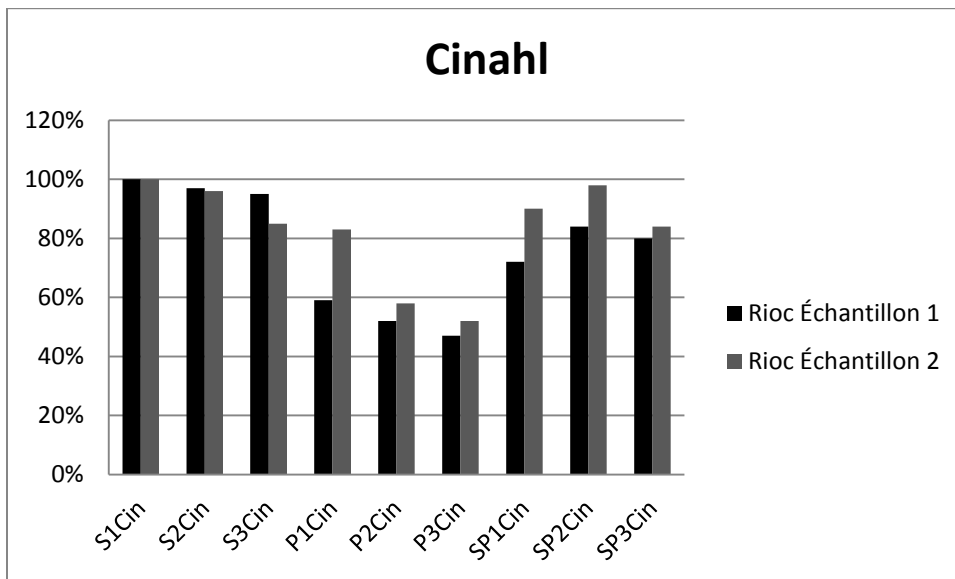


Tableau 24 : Statistiques RIOC pour les filtres de recherche en utilisant les échantillons 1 et 2.

Filtre de recherche	Échantillon 1			Échantillon 2			Valeur de la Statistique du test de comparaison des RIOCs
	RIOC	95% CI	(RIOC SRIOC*)	RIOC	95% CI	(RIOC SRIOC*)	
Medline							
<i>S1Med</i>	100%	[1,000; 1,000]	67	100%	[1,000; 1,000]	89	---
<i>S2Med</i>	99%	[0,957; 1,014]	447	95%	[0,892; 1,006]	673	1,14
<i>S3Med</i>	99%	[0,956; 1,014]	388	95%	[0,889; 1,007]	520	1,14
<i>P1Med</i>	49%	[0,377; 0,600]	377	74%	[0,631; 0,853]	1146	-3,21
<i>P2Med</i>	67%	[0,558; 0,772]	578	49%	[0,361; 0,615]	655	2,14
<i>P3Med</i>	58%	[0,466; 0,689]	499	41%	[0,286; 0,536]	591	2,01
<i>SP1Med</i>	92%	[0,857; 0,982]	677	95%	[0,896; 1,006]	1279	-0,74
<i>SP2Med</i>	88%	[0,805; 0,955]	649	93%	[0,871; 0,998]	1257	-1,11
<i>SP3Med</i>	95%	[0,893; 0,998]	642	95%	[0,895; 1,006]	1140	-0,12
Embase							
<i>S1Emb</i>	100%	[1,000; 1,000]	378	94%	[0,882; 1,007]	504	1,78
<i>S2Emb</i>	98%	[0,954; 1,015]	375	94%	[0,882; 1,007]	510	1,15
<i>S3Emb</i>	94%	[0,876; 0,997]	320	92%	[0,852; 0,997]	413	0,26
<i>P1Emb</i>	63%	[0,514; 0,741]	388	52%	[0,391; 0,653]	695	1,21
<i>P2Emb</i>	64%	[0,527; 0,753]	389	52%	[0,391; 0,653]	695	1,36
<i>P3Emb</i>	65%	[0,539; 0,764]	379	54%	[0,409; 0,670]	765	1,31
<i>SP1Emb</i>	77%	[0,673; 0,872]	435	81%	[0,705; 0,913]	949	-0,50
<i>SP2Emb</i>	79%	[0,687; 0,883]	424	88%	[0,790; 0,964]	884	-1,40
<i>SP3Emb</i>	85%	[0,768; 0,939]	408	84%	[0,747; 0,939]	949	0,16
CINAHL							
<i>S1Cin</i>	100%	[1,000; 1,000]	133	100%	[1,000; 1,000]	354	----
<i>S2Cin</i>	97%	[0,921; 1,025]	120	96%	[0,894; 1,017]	304	0,44
<i>S3Cin</i>	95%	[0,877; 1,018]	123	85%	[0,741; 0,954]	294	1,56
<i>P1Cin</i>	59%	[0,501; 0,681]	77	83%	[0,749; 0,909]	165	-3,95
<i>P2Cin</i>	52%	[0,390; 0,653]	190	58%	[0,450; 0,717]	441	-0,66
<i>P3Cin</i>	47%	[0,340; 0,597]	154	52%	[0,382; 0,650]	388	-0,52
<i>SP1Cin</i>	72%	[0,595; 0,847]	286	90%	[0,812; 0,985]	711	-2,31
<i>SP2Cin</i>	84%	[0,727; 0,941]	269	98%	[0,938; 1,020]	668	-2,54
<i>SP3Cin</i>	80%	[0,678; 0,910]	263	84%	[0,732; 0,943]	635	-0,55

L'analyse de la performance des filtres dans le gold standard NPSF.

La capacité des filtres de recherche à repérer des articles du gold standard NPSF dans la base de données Medline.

Pour toutes les stratégies de Medline, la sensibilité diminue et la précision augmente, ce qui confirme que les stratégies que nous avons élaborées ne couvrent pas de façon adéquate le champ de la sécurité des patients tel que défini par la National Patient Safety Foundation. Les stratégies de Medline ont perdu en moyenne 38% de leur sensibilité. Ce chiffre est de 29,3% pour les stratégies les plus sensibles, de 36,7% pour les stratégies les plus précises et de 48,2% pour les stratégies les plus équilibrées en termes de sensibilité et de précision. La stratégie la plus sensible de Medline permet de repérer plus de 85% du matériel du gold standard NPSF, mais au prix d'un grand effort en matière de documents non pertinents à trier.

Elles n'en sont pas moins utiles pour autant. Dans une perspective de revue systématique et en l'absence de filtres plus adaptés pour couvrir le champ de la sécurité des patients tel que défini par le gold standard NPSF, les stratégies S3Med et S2Med de même que les stratégies SP1Med, SP2Med et SP3Med semblent être les plus avantageuses pour débiter une recherche bibliographique car elles offrent un bon équilibre entre la sensibilité et l'effort à fournir.

Les stratégies P1Med et P2Med restent utiles pour les utilisateurs qui souhaitent maximiser la précision, car leur précision déjà bonne dans les deux premiers échantillons qui ont servi à leur construction et leur évaluation est encore meilleure dans le gold standard NPSF.

Tableau 25 : Récapitulatif de la performance des filtres de Medline dans les 3 échantillons.

		Sensibilité	Précision ¹	Spécificité	Sensibilité * Précision ¹	Nombre d'articles à lire ¹
S1Med	Échantillon 1	100,00%	8,28%	45,56%	8,28%	12
	Échantillon 2	100%	3,51%	59,36%	3,51%	28
	Échantillon 3	85,9%	17%	Inconnu	14,60%	6
S2Med	Échantillon 1	98,80%	29,08%	88,17%	28,73%	3
	Échantillon 2	95,31%	17,63%	93,41%	16,80%	6
	Échantillon 3	64,60%	48,00%	Inconnu	31,01%	2
S3Med	Échantillon 1	98,80%	25,87%	86,09%	25,56%	4
	Échantillon 2	95,31%	13,96%	91,31%	13,30%	2
	Échantillon 3	59,40%	56,00%	Inconnu	33,26%	2
P1Med	Échantillon 1	45,78%	51,35%	97,87%	23,51%	2
	Échantillon 2	75%	36,64%	98,08%	27,48%	3
	Échantillon 3	24,20%	91,00%	Inconnu	22,02%	1
P2Med	Échantillon 1	68,67%	51,35%	96,80%	35,27%	2
	Échantillon 2	50%	32%	98,43%	16%	3
	Échantillon 3	26,80%	90,00%	Inconnu	24,12%	1
P3Med	Échantillon 1	60,24%	51,02%	97,16%	30,74%	2
	Échantillon 2	42,19%	34,18%	98,80%	14,42%	3
	Échantillon 3	13,60%	41,00%	Inconnu	5,58%	2
SP1Med	Échantillon 1	92,77%	44,25%	94,26%	41,05%	2
	Échantillon 2	95,31%	32,111%	97,02%	30,60%	3
	Échantillon 3	42,90%	72,00%	Inconnu	30,89%	1
SP2Med	Échantillon 1	89,16%	44,31%	94,50%	39,51%	2
	Échantillon 2	93,75%	32,09%	97,07%	30,08%	3
	Échantillon 3	41,80%	74,00%	Inconnu	30,93%	1
SP3Med	Échantillon 1	95,18%	41,15%	93,31%	39,16%	2
	Échantillon 2	95,31%	28,77%	96,51%	27,42%	3
	Échantillon 3	47,80%	67,00%	Inconnu	32,03%	1

Échantillon 1: N = 1773; gold standard = 83
Échantillon 2: N = 4392; gold standard = 64
Échantillon 3: N = Inconnu; NPSF gold standard = 3254
¹ pour l'échantillon 3, les valeurs de la précision, du produit de la Sensibilité par la Précision et du nombre d'articles à lire sont les valeurs estimées à partir du calcul de
 $Précision_{Est}, Sensibilité * Précision_{Est}$ et $Nombre\ d'articles\ à\ lire_{Est}$

La capacité des filtres de recherche à repérer des articles du gold standard NPSF dans la base de données Embase.

Contrairement aux stratégies élaborées dans Medline, les stratégies élaborées pour Embase reflètent mieux le critère « organisation de santé » qui avait été retenu pour guider le choix des gold standard dans l'échantillon 1 et 2. Dans le gold standard NPSF qui ne se limite pas aux organisations de santé, ces stratégies sont toutes moins sensibles. Les stratégies d'Embase ont perdu en moyenne 42% de leur sensibilité. Ce chiffre est de 46,5% pour les stratégies les plus sensibles, de 35,9% pour les stratégies les plus précises et de 43,6% pour les stratégies les plus équilibrées en termes de sensibilité et de précision.

La précision augmente également, mais un peu moins qu'avec les stratégies de Medline.

Les stratégies les plus sensibles et les plus équilibrées d'Embase offrent toutes un bon équilibre entre sensibilité et effort. En l'absence de stratégies plus adaptées pour couvrir le champ de la Sécurité des Patients tel que défini par le gold standard NPSF, elles offrent un bon point de départ pour une revue systématique. Les stratégies P1Emb et P2Emb restent pertinentes pour les personnes qui désirent maximiser la précision.

Tableau 26 : Récapitulatif de la performance des filtres d'Embase dans les 3 échantillons.

		Sensibilité	Précision ¹	Spécificité	Sensibilité * Précision ¹	Nombre d'articles à lire ¹
S1Emb	Échantillon 1	100.00%	30.04%	86.60%	30.04%	3
	Échantillon 2	95.00%	19.52%	92.13%	18.54%	5
	Échantillon 3	42,70%	61,00%	Inconnu	26,05%	2
S2Emb	Échantillon 1	98.73%	30.23%	86.89%	29.85%	3
	Échantillon 2	95.00%	19.72%	92.23%	18.74%	5
	Échantillon 3	56,60%	53,00%	Inconnu	30,00%	2
S3Emb	Échantillon 1	94.94%	27.68%	85.72%	26.27%	4
	Échantillon 2	93.33%	16.67%	90.62%	15.56%]	6
	Échantillon 3	54,80%	45,00%	Inconnu	24,66%	2
P1Emb	Échantillon 1	65.82%	45.61%	95.48%	30.02%	2
	Échantillon 2	53.33%	45.71%	98.73%	24.38%	2
	Échantillon 3	33,10%	72,00%	Inconnu	23,83%	1
P2Emb	Échantillon 1	67.09%	44.92%	95.27%	30.13%	2
	Échantillon 2	53.33%	45.71%	98.73%	24.38%	2
	Échantillon 3	33,30%	64,00%	Inconnu	21,31%	2
P3Emb	Échantillon 1	68.35%	43.20%	94.83%	29.53%	2
	Échantillon 2	55%	48.53%	98.83%	26.69%	2
	Échantillon 3	27,20%	38,00%	Inconnu	10,34%	3
SP1Emb	Échantillon 1	79.75%	42.00%	93.66%	33.49%	2
	Échantillon 2	81.67%	40.50%	97.59%	33.07%	2
	Échantillon 3	40,50%	45,00%	Inconnu	18,23%	2
SP2Emb	Échantillon 1	81.01%	40.51%	93.15%	32.82%	2
	Échantillon 2	85.00%	38.93%	97.32%	33.09%	3
	Échantillon 3	35,60%	36,00%	Inconnu	12,82%	3
SP3Emb	Échantillon 1	87.34%	36.51%	91.26%	31.89%	3
	Échantillon 2	88.33%	35.10%	96.72%	31.00%	3
	Échantillon 3	41,20%	33,00%	Inconnu	13,60%	3

Échantillon 1: N = 1452 gold standard = 79
Échantillon 2: N = 3046 gold standard = 61
Échantillon 3: N = Inconnu NPSF gold standard = 2294
¹ pour l'échantillon 3, les valeurs de la précision, du produit de la Sensibilité par la Précision et du nombre d'articles à lire sont les valeurs estimées à partir du calcul de
 $Précision_{Est}, Sensibilité * Précision_{Est}$ et $Nombre\ d'articles\ à\ lire_{Est}$

La capacité des filtres de recherche à repérer des articles du gold standard NPSF dans la base de données CINAHL.

Tout comme les stratégies élaborées pour Medline et pour Embase, les stratégies élaborées pour CINAHL perdent en sensibilité dans le gold standard NPSF. Les stratégies de CINAHL ont perdu en moyenne 33,8% de leur sensibilité. Ces chiffres sont de 40,8% pour les stratégies les plus sensibles, de 21,26% pour les stratégies les plus précises et de 39,35% pour les stratégies les plus équilibrées. Par contre cette perte en sensibilité n'est pas compensée par une augmentation de la précision comme c'est le cas dans les autres bases de données à l'exception des stratégies SP1Cin et SP3Cin. Dans la perspective d'une revue systématique et en l'absence de stratégies plus appropriées pour repérer des articles portant sur la sécurité des patients répondant aux critères exhibés dans le gold standard NPSF, la stratégie SP1Cin est celle qui offre le meilleur rapport entre sensibilité et effort de recherche, suivie par la stratégie SP3Cin et les trois stratégies les plus sensibles S1Cin, S2Cin et S3Cin. Les personnes désireuses de maximiser la précision pourront utiliser la stratégie P3Cin.

Tableau 27 : Récapitulatif de la performance des filtres de CINAHL dans les 3 échantillons.

		Sensibilité	Précision ¹	Spécificité	Sensibilité * Précision ¹	Nombre d'articles à lire ¹
S1Cin	Échantillon 1	100.00%	18.30%	73.23%	18.30%	5
	Échantillon 2	100%	18.71%	89.35%	18.71%	5
	Échantillon 3	60,60%	20,00%	Inconnu	12,12%	5
S2Cin	Échantillon 1	98.21%	17.41%	72.06%	17.09%	6
	Échantillon 2	96.15%	17.06%	88.55%	16.41%	6
	Échantillon 3	60,80%	19,00%	Inconnu	11,55%	5
S3Cin	Échantillon 1	96.43%	18.06%	73.77%	17.42%	6
	Échantillon 2	86.54%	18.37%	90.58%	15.89%	5
	Échantillon 3	51,00%	32,00%	Inconnu	16,32%	3
P1Cin	Échantillon 1	14.29%	61.54%	99.46%	8.79%	2
	Échantillon 2	9.62%	83.33%	99.95%	8.01%	1
	Échantillon 3	2,30%	22,00%	Inconnu	0,51%	5
P2Cin	Échantillon 1	39.29%	55.00%	98.07%	21.61%	2
	Échantillon 2	36.54%	59.38%	99.39%	19.28%	2
	Échantillon 3	11,20%	32,00%	Inconnu	3,58%	3
P3Cin	Échantillon 1	35.71%	50.00%	97.86%	17.86%	2
	Échantillon 2	36.54%	52.78%	99.20%	19.28%	2
	Échantillon 3	12,00%	50,00%	Inconnu	6,00%	2
SP1Cin	Échantillon 1	75.00%	43.30%	94.11%	32.47%	2
	Échantillon 2	90.38%	38.84%	96.51%	35.11%	3
	Échantillon 3	42,20%	100,00%	Inconnu	42,20%	1
SP2Cin	Échantillon 1	85.71%	36.36%	91.01%	31.17%	3
	Échantillon 2	98.08%	33.77%	95.29%	33.13%	3
	Échantillon 3	45,40%	28,00%	Inconnu	12,71%	4
SP3Cin	Échantillon 1	82.14%	37.10%	91.65%	30.47%	3
	Échantillon 2	84.62%	37.29%	96.51%	31.55%	3
	Échantillon 3	37,20%	46,00%	Inconnu	17,11%	2

Échantillon 1: N = 990 gold standard = 56

Échantillon 2: N = 2175 gold standard = 52

Échantillon 3: N = Inconnu NPSF gold standard = 1895

¹ pour l'échantillon 3, les valeurs de la précision, du produit de la Sensibilité par la Précision et du nombre d'articles à lire sont les valeurs estimées à partir du calcul de

$Précision_{Est}, Sensibilité * Précision_{Est}$ et $Nombre\ d'articles\ à\ lire_{Est}$

Discussion

L'évaluation de la qualité des décisions prises à l'aide des filtres de recherche confirme leur utilité pour repérer des articles portant sur la sécurité des patients dans les organisations de santé, explicitement inscrits dans le champ de la sécurité des patients par leurs auteurs. Pour certains des filtres de recherche, notamment P1Med, P2Med, P3Med, P1Cin, SP1Cin et SP2Cin, le pourcentage d'amélioration de la prise de décision que permet le filtre de recherche comparativement à une décision basée sur un processus purement aléatoire varie significativement entre les deux échantillons dans lesquels l'utilité a été évaluée. Ces résultats complètent ceux de l'évaluation de la validité de critères des filtres de recherche à l'aide du coefficient de validité, notamment avec la possibilité qui est offerte de tester les différences entre les valeurs de la statistique RIOC : les stratégies P1Med, P1Cin, SP1Cin et SP2Cin représentent mieux le critère dans l'échantillon 2 et les stratégies P2Med, et P3Med représentent mieux le critère dans l'échantillon 1. Pour les autres filtres de recherche, l'utilité est la même dans les deux échantillons.

Nous avons démontré que ces deux échantillons étaient significativement différents non seulement en termes de proportions d'articles définis comme faisant partie du gold standard, mais aussi en termes de proportions d'articles repérés par les filtres de recherche. Cependant l'utilité des filtres n'est pas nécessairement plus élevée dans l'échantillon 1 qui a servi à la construction des filtres de recherche que dans l'échantillon 2.

Lorsque l'on change de critère et que le gold standard comprend des articles sélectionnés selon une perspective plus large de la sécurité des patients, les filtres que nous proposons deviennent de façon générale moins sensibles, un phénomène souligné par de nombreux auteurs ([Haynes, Wilczynski et al. 1994](#); [White, Glanville et al. 2001](#); [Vincent, Greenley et al. 2003](#)).

Cette baisse de sensibilité n'est pas surprenante car les critères que nous avons retenus pour la construction des filtres représentent un sous ensemble du champ de la

sécurité des patients tel que définis par le gold standard NPSF car le gold standard NPSF contient non seulement des articles qui traitent de la sécurité des patients et de la qualité des soins, mais aussi des articles qui traitent de la qualité des soins sans aucune référence explicite à la sécurité des patients. Pour de nombreux experts, la sécurité des patients est une sous-catégorie de la qualité des soins ([Emanuel, Berwick et al. 2008](#)), ce qui pourrait laisser penser que tout ce qui est sécurité des patients est qualité des soins, mais que tout ce qui est qualité des soins n'est pas nécessairement sécurité des patients. Nous avons fait le choix de limiter notre gold standard à des articles qui traitent de façon explicite de la sécurité des patients, sachant que certains vont également porter sur la qualité des soins afin d'éviter le deuxième cas de figure. En outre, le gold standard NPSF ne se limite pas de façon exclusive aux organisations de santé.

Compte tenu des différences importantes qui existent entre les gold standards expérimentaux et le gold standard NPSF, les performances obtenues surtout par les stratégies les plus sensibles et les stratégies les plus équilibrées, qui sont aussi les stratégies les plus valides pour repérer des articles répondant à nos critères, sont assez remarquables.

Conclusion

La comparaison avec le gold standard NPSF illustre à quel point le champ de la sécurité des patients est large et la nécessité d'élaborer des filtres de recherche qui puissent couvrir adéquatement son contenu car l'information constitue un allié important dans la quête d'une amélioration de la sécurité des patients ([Huehns and Fletcher 2010](#)).

Les filtres de recherche que nous proposons sont des outils performants et valides pour repérer les articles portant sur la sécurité des patients dans les organisations de santé qui ont été explicitement inscrits dans le champ de la sécurité des patients par leurs auteurs. Ils ne couvrent pas tout le champ de la sécurité des patients, mais offrent une bonne alternative pour initier une revue systématique, même si la définition de la sécurité des

patients utilisée est plus large que celles que nous avons adoptée lors de la construction des filtres. Leur utilisation ne devrait pas exclure d'autres moyens d'avoir accès à la littérature. Elle peut par exemple être complétée par une revue manuelle qui reste un outil important dans le cadre d'une revue systématique ([Hopewell, Clarke et al. 2007](#)), par la recherche de la littérature grise qui n'est pas toujours bien couverte dans les bases de données bibliographiques ([McAuley, Pham et al. 2000](#); [Avenell, Handoll et al. 2001](#); [Egger, Juni et al. 2003](#); [Benzies, Premji et al. 2006](#)), et par des recherches dans d'autres bases de données car en raison de la grande dispersion de l'information dans le champ de la santé, il est important de combiner différentes bases de données pour faire une revue systématique ([Helmer, Savoie et al. 2001](#); [Bickley and Harrison 2003](#); [Kelly and St Pierre-Hansen 2008](#); [Lawrence 2008](#)).

Ultimement il appartient aux utilisateurs de faire le jugement final sur la pertinence des documents qui auront été repérés par les filtres de recherche bibliographique. Ce jugement va dépendre de leur connaissance du domaine, des objectifs de leur recherche d'information, de l'étape dans laquelle ils sont dans leur recherche, etc. ([Su 1993](#); [Ruthven 2005](#); [Saracevic 2007](#); [Saracevic 2008](#)).

Conflits d'intérêts

Les auteurs déclarent qu'ils n'ont pas d'intérêts concurrents.

Contributions

Anaïs Tanon, André-Pierre Contandriopoulos and François Champagne ont contribué à la conception et au design de l'étude. Affaud Anaïs Tanon était responsable de l'acquisition, de l'analyse et de l'interprétation des données. Tous les auteurs ont contribué à la révision critique de l'article avant sa soumission.

Remerciements

Cette recherche a été financée par les Instituts de Recherche en Santé du Canada (IRSC). Des bourses ont été offertes par : le programme AnÉIS (*Analyse et Évaluation des Interventions en Santé*), par le biais d'un projet de recherche financé par les IRSC (*Instituts de Recherche en Santé du Canada*), par le CRSH (*Conseil de Recherche en Sciences Humaines du Canada*), par le GIRU (*Groupe de recherche interuniversitaire sur les urgences*); par le gouvernement de la Côte d'Ivoire, par l'Université de Montréal.

Le projet a été mené avec le soutien du Programme de Takemi en Santé Internationale à l'école de Santé Publique de l'Université Harvard. Les auteurs remercient le professeur Michelle Mello et le professeur Ashish Jha, mentors de Anaïs Tanon à l'Université Harvard.

Chapitre 8 : Un guide pour la construction, l'évaluation et la validation des filtres de recherche bibliographique.

Paper 5: Retrieving scientific papers in bibliographic databases: a step-by-step guide to validated search filters construction

Affaud Anaïs Tanon, François Champagne, André-Pierre Contandriopoulos

Abstract

Health care practitioners, decision-makers, and researchers need to acquire the skills necessary to identify their information needs and to better locate, evaluate, and effectively use that information to take advantage of the opportunities presented by the latest technological developments in the information and communications sectors. One of these skills is the ability to develop search filters to retrieve papers that meet specific information needs in bibliographic databases. Despite the fact that more and more resources dealing with the methodology of validated search filters development are available in the literature, there is, to our knowledge, no compact guide for the development and validation of bibliographic search filters. Therefore, we propose a guide, based on our experience and optimized through a literature review, to assist health researchers, policy-makers, and practitioners in their information retrieval activities.

Keywords: information storage and retrieval; evidence-based medicine; bibliographic database; delivery of health care; public health

Introduction

Health care practitioners, decision-makers, and researchers need to acquire the skills necessary to identify their information needs and to better locate, evaluate, and effectively use that information to take advantage of the opportunities presented by the latest technological developments in the information and communications sectors, and to meet the challenges of health care in the early 21st century ([Snavely and Cooper 1997](#); [Harley 2001](#); [Gebbie, Rosenstock et al. 2003](#); [Saranto and Hovenga 2004](#); [Albitz 2007](#); [Cobus 2008](#)). Although health information can be found in several sources, bibliographic databases are almost always most likely to contain the majority of scientific evidence ([Haig and Dozier 2003a](#)). However, people often experience retrieving information from that source as a difficult and frustrating process because they feel overwhelmed by the amount of information obtained and unsatisfied with the results of the information search ([Davies 2007](#); [Hemminger, Lu et al. 2007](#); [Jackson, Baird et al. 2007](#); [Bawden and Robinson 2009](#); [MacDonald, Bath et al. 2011](#)).

For more than 40 years, research in the information retrieval field has been dedicated to improving and creating ways to support people in better accessing the information they need.

To achieve this goal in the health and biomedical fields, solutions to improve information retrieval in bibliographic databases usually follow one of two paths: a) well established techniques in the information retrieval field are applied to significant real world problems, or b) new ideas about information retrieval systems or features are conceptualized, evaluated through experiments, and transformed into new applications ([Järvelin 2011](#)).

Consequently, many tools based on advanced information retrieval features are now available to the health community. For example, Pubmed is a web-based interface to the very well-known Medline database that allows the automatic expansion of queries entered by users by including additional synonymous terminology to improve retrieval performance ([Lu, Kim et al. 2009](#)). AliBaba is a web-based tool that searches protein and disease

centered information on Pubmed and then displays that information in graphs showing biological or medical objects such as proteins, diseases, or drugs as nodes and meaningful associations between them as edges ([Plake, Schiemann et al. 2006](#)). GoPubMed is a web search engine that has access to the Medline database and presents the results of research on genes or proteins in hierarchical clusters; scientific papers are presented to the user in groups by subject using Medline indexing MeSH terms and a gene ontology ([Doms and Schroeder 2005](#)). Other examples of innovative applications in the biomedical field are described by [Kim and Rebholz-Schuhmann \(2008\)](#).

Standard techniques for optimizing retrieval tasks have also been applied to retrieving information in health bibliographic databases. In these tasks, people specify the information they are looking for through a query, which initiates a search executed by the information retrieval system for documents likely to be relevant to the user. These techniques, established through experiments conducted by Cyril W. Cleverdon at Cranfield University in the 1960s for systems displaying results in an unordered fashion ([Cleverdon and Keen 1966](#); [Harter and Hert 1997](#); [Kagolovsky and Moehr 2003](#)), and subsequently by experiments like TREC (Text REtrieval Conference) ([Voorhees 2009](#)) for ranked information retrieval systems, are well-known in the information retrieval community. They gained popularity in the health field as part of the evidence-based medicine movement ([Sackett, Straus et al. 2000](#)).

In an environment where hundreds of resources are added every day to those already available in health bibliographic databases ([Hunter and Cohen 2006](#); [U.S. National Library of Medicine - National Institutes of Health 2011](#)), optimized search filters (queries) can help locate information more efficiently ([Ford 1999](#); [Kramer 2011](#)).

The death of a volunteer in a clinical trial at the Johns Hopkins Hospital in 2001 ([Ramsay 2001](#)) is a tragic reminder of the importance of the information retrieval step in the evidence-based decision making process. Following the investigation conducted to clarify the circumstances of this death, the inadequacy of the literature search used to build the research protocol was emphasized ([Steinbrook 2002](#)).

Optimized search filters can also help reduce the time needed to produce systematic reviews ([Sampson, Shojania et al. 2008](#); [Major, Warren et al. 2009](#)).

Optimized bibliographic search filters are logical combinations of text terms (words found in the title, abstract, or body of a document) and indexing terms (controlled vocabulary of a database) designed to identify studies meeting specific methodological criteria or covering a specific topic in a bibliographic database. They are called “optimized search filters” because more objective approaches are involved in their design ([Jenkins 2004](#)).

Optimized search filters first appeared in the published health literature in the mid-1980s ([Bak, Mierzewski-Urban et al. 2009](#)). Since then, they have continued to grow in popularity ([Jenkins and Johnson 2004](#); [Glanville, Bayliss et al. 2008](#)). As a result, a growing number of optimized search filters are available in the scientific health literature for different types of studies, topics, and bibliographic databases ([Jenkins 2004](#); [Glanville, Bayliss et al. 2008](#))

Despite this growth, it could be necessary to develop optimized custom filters that meet the specific needs of a given project. With this in mind, we propose a guide for developing optimized search filters.

Method

To prepare this document we built upon our experience with constructing and validating search filters to retrieve patient safety papers on classic Boolean information retrieval systems such as Ovid Medline, Embase and CINAHL ([Tanon, Champagne et al. 2010](#)). Subsequently, we conducted a series of literature searches using the pearl growing and citation searching methods to identify articles that allowed us to address the issues raised at each stage. The pearl growing method identifies items by beginning with one article considered relevant to the topic and then analyzing how this article was indexed in the databases considered ([Schlosser, Wendt et al. 2006](#)). The citation searching method can

be used to identify items from the references cited in the relevant section (background reference) and references that cite the relevant section (forward citation) ([Hammond and Brown 2008](#)).

The searches were conducted in the Medline, Web of Science, LISA and LISTA databases. The different steps of the guide are presented in Table 28 and described in greater detail in the following sections.

Table 28. Steps for search filter development, assessment and validation

	Steps
Search filter development and assessment	(1) Formulate the question(s) (2) Define the inclusion and exclusion criteria (3) Plan the evaluation of the search filters' performance (4) Select the keywords (5) Evaluate and combine the individual search terms
Search filter validation	(6) Validation of the search filter (<i>criterion-related</i>)

Formulate the question(s)

Developing a good research question is one of the most important steps in any research or decision-making evidence-based process ([Lipowski 2008](#)).

As part of the development of search filters, this first step:

- defines the different facets of the research to guide the identification of keywords, in accordance with the literature search strategy called the “building block approach” ([Harter 1986](#));
- facilitates the choice of selection criteria for inclusion and exclusion;
- guides the identification and selection of keywords that will be included in the search filters;

- guides the combination of the keywords.

The literature is full of tips to help formulate good research questions. In clinical research, for example, the PICO format has been proposed to guide the development of good questions ([Corrall, Wyer et al. 2002](#); [Melnyk and Fineout-Overholt 2002](#); [Bragge 2010](#); [Burns and Chung 2010](#)). Questions in this format are usually in the form: “In [population], what is the effect of [intervention] on [outcome] compared with [comparison intervention]?” ([Bragge 2010](#)).

Under this format, the questions should (a) be specific, (b) identify the population and the problem, (c) define the procedure or area of interest, (d) identify interventions being compared, if applicable, and (e) refer to measurable results.

Developing good questions requires a preliminary review of the literature in order to avoid asking inappropriate questions or focusing on issues that have already been discussed ([Bragge 2010](#)). This preliminary review need not be exhaustive. However, it must be of sufficient magnitude to allow a certain familiarity with the problem studied and knowledge of the latest developments on the issue ([Haynes 2006](#)).

Define the inclusion and exclusion criteria

Defining the inclusion and exclusion criteria is another extremely important step in the development of search filters ([Petrosino 1995](#)). The criteria will be used to make a binary judgment (yes or no) about relevance based on the articles’ content to select those that will serve as references for constructing the search filters. The inclusion and exclusion criteria must derive from the selected research question and will be developed from a preliminary review of the literature.

The choice of inclusion criteria will determine the extent to which the search filters developed will be valid instruments capable of retrieving other documents sharing the methodology or topic outside the experimental context in which the filters have been

developed. It is therefore important that the inclusion and exclusion criteria be clearly defined from the beginning of the project so that users will know the specific contexts in which the search filters can be applied. In addition, explicit criteria allow replication and updating of the search filters, which is especially important in a field of research that is constantly evolving, such as the patient safety field.

It might be also useful at this stage to review the inclusion and exclusion criteria that have been applied to similar studies, if available in the literature.

Plan the evaluation of search filters' performance

Indicators for evaluating search filters

The indicators used to evaluate search filters' performance mirror those used in epidemiological test development ([Jenkins 2004](#); [Straus 2005](#); [Manning, Raghavan et al. 2008](#)), as shown in Table 29, where: **a** is the number of relevant documents retrieved; **b** is the number of irrelevant documents retrieved; **c** is the number of relevant documents not retrieved; **d** is the number of irrelevant documents not retrieved; **(a+b)** is the set of retrieved articles; **(c+d)** is the set of articles not retrieved; **(a+c)** is the set of relevant articles; and **(b+d)** is the set of irrelevant articles.

Sensitivity and precision are the two indicators most often used to compare search filter performance ([Hersh 2009](#)). Sensitivity is the proportion of relevant documents retrieved by the search filters, while precision is the proportion of documents retrieved by the search filters that is relevant. These two indicators move in opposite directions: the higher the sensitivity of a search filter, the lower the precision ([Buckland and Gey 1994](#)). Specificity is the complement of sensitivity. It measures the proportion of irrelevant documents that are not retrieved by the search filters. The number of articles to read (Number needed to read) is the inverse of precision. This indicator provides information on the number of items one needs to read before finding a relevant article retrieved by the

search filter ([Bachmann, Coray et al. 2002](#)) The product of sensitivity and precision is another indicator that offers a trade-off between sensitivity and precision. Many other indicators are available to evaluate the performance of search filters in unordered collections of papers ([Sanderson 2010](#)). Table 30 shows some examples of indicators.

Table 29: Classification matrix for the evaluation of search filters

		Set of papers		Total
		Relevant (gold standard) a (True positives)	Non relevant b (False positives)	
Search filter	Papers retrieved	a	b	(a + b)
	Papers not retrieved	c (False negatives)	d (True negatives)	(c+d)
Total		(a+c)	(b+d)	N = a+b+c+d

Table 30: Indicators for evaluating search filters

Evaluation indicators	Epidemiological test development
Sensitivity (recall): $a/(a + c)$	True positive rate (Sensitivity, Hit rate, Recall): $TP/(TP + FN)$
Silence (1-sensitivity): $c/(a + c)$	False negative rate: $FN/(TP + FN)$
Precision: $a/(a + b)$	Positive predictive value (Precision): $TP/(TP + FP)$
Noise (1-Precision) : $b/(a + b)$	False discovery rate: $FP/(TP + FP)$
Fall-out: $b/(b + d)$	False positive rate (False alarm rate): $FP/(FP + TN)$
Specificity: $d/(b + d)$	Specificity (True negative rate): $TN/(FP + TN)$
Number needed to read: $(a + b)/a$	Number needed to test: $(TP + FP)/TP$

Collect a set of papers

To measure the performance of the search filters, a set of documents must be defined. It is composed of two subsets: the relevant papers that should be retrieved with the search filters (the gold standard), and the irrelevant papers that should not be retrieved using the search filters.

Select the journals

The exercise begins with the selection of journals in which articles belonging to the gold standard will be identified. Several approaches can be used. The search can be done directly:

- using periodicals directories such as Ulrich's Periodicals Directory⁷;
- using available lists of papers relevant to the subject of interest.

The search can also be done indirectly by deriving a list of journals from a collection of relevant articles. This approach can be started:

- with a large and known bibliographic database on the subject. For example, the National Patient Safety Foundation publishes an annotated bibliography of publications on patient safety for its members (Current Awareness Literature Alert)⁸;

⁷ Ulrich is a directory that indexes more than 300,000 periodicals published throughout the world (with particular emphasis on the English language). For each periodical, ULRICH provides information on: ISSN; title and previous titles; starting date; place of publication and publisher; cost; availability of electronic versions; subscription terms; approximate circulation as estimated by the publisher; subject information, searchable as subject terms or approximate Dewey Classification; special features; indexing information; whether the publication is available on open access; and whether the publication is peer-reviewed, which is taken to include professional magazines with equivalent editorial control of quality.

⁸ <http://www.npsf.org/for-healthcare-professionals/resource-center/npsf-publications-members-only/> (accessed Dec 2011).

- with the bibliographies of relevant articles to build a custom collection ([Sampson, Zhang et al. 2006](#); [Wilczynski, Garg et al. 2007](#));
- with an electronic search in a bibliographic database using simple or more complex strategies to identify a preliminary set of relevant articles ([Hopewell, Clarke et al. 2007](#)).

Once a list of journals is available, the final selection must be made taking into account factors such as the number of journals selected, the type of journals selected, publication languages, disciplines covered, etc., as well as other elements that could undermine the validity of search filters ([Jenkins 2004](#)). Sampling strategies should provide an adequate number of gold standard papers, as well as support validation requirements.

Select the databases

Depending on the objectives of the research, the decision can be made to develop search filters for a single bibliographic database or for multiple databases.

If resources permit, it is best to develop filters for multiple databases, because the information produced in the health care field is scattered across many disciplines and therefore throughout a large number of bibliographic resources ([Hood and Wilson 2001](#); [Haig and Dozier 2003a](#); [McKibbon, Wilczynski et al. 2004](#); [Betrn, Say et al. 2005](#); [Crumley, Wiebe et al. 2005](#); [Lawrence 2008](#); [Whiting, Westwood et al. 2008](#)) It is also often published outside the channels that might be described as “natural”. A bibliometric study showed that nearly half of relevant articles in renal medicine are published outside of the journals of the specialty ([Garg, Iansavichus et al. 2006](#)). Another study found that nearly 80% of articles on obesity are found outside the journals of the specialty ([Baier, Wilczynski et al. 2010](#)). In addition, for the same database, the results of the search may be different depending on the selected interface ([Younger and Boddy 2009](#)).

To choose the databases to be used for constructing filters, one can start either with the objectives of the research, and base the selection of the databases on the type of information sought ([Wyer, Allen et al. 2003](#)), or with the list of journals used to select the gold standard papers, to identify the databases in which these journals are indexed ([Schlosser, Wendt et al. 2006](#)).

Finally, the choice of databases should also be based on factors such as ease of access to the database, publication type, language and the time period covered, the availability of full text, the presence of an indexing system, etc. ([Hart 2001](#)), as well as on the structure of the database.

Selecting the gold standard papers

Identification of papers to include in the gold standard can be done in two ways:

- Manually: hard copies of all papers selected are read and the inclusion and exclusion criteria are applied to split the group of papers between the set of relevant articles (gold standard) and the set of non-relevant articles ([Delamere and Williams 2001](#));
- Electronically: articles published in the selected journals are identified in the chosen databases. The full text of articles deemed relevant based on title, abstract, and indexing keywords are read to make a final selection based on the inclusion and exclusion criteria.

The manual method is inherently more laborious but gives better results in the identification of relevant articles not indexed in bibliographic databases ([Hopewell, Clarke et al. 2007](#)). The electronic method is still relevant, however, because ultimately in developing search filters we are interested in articles indexed in selected databases.

Technological interfaces such as the SRS platform for systematic reviews, TrialStat, effectively automate many tasks necessary for identification of the gold standard from an

electronic search ([O'Blenis 2004](#)). Other tools are available ([Bax, Yu et al. 2007](#); [Littell, Corcoran et al. 2008](#)), such as RevMan (Reviewer Manager), the software used for preparing and maintaining Cochrane Reviews ([The Cochrane Collaboration 2011](#)), or EPPI Reviewer 4, developed by the Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre), which is part of the Social Science Research Unit at the Institute of Education, University of London ([Thomas, Brunton et al. 2010](#)).

Whether the gold standard is identified manually or electronically, it is important to document, for all articles examined, the reasons for either inclusion or exclusion.

Select the keywords

The keywords are composed of indexing terms (a controlled vocabulary available in some databases such as Medline) and text terms (the vocabulary found, for example, in the articles' titles and abstracts). It is preferable to combine these two types of search terms, because while indexing terms are generally more sensitive, more specific, and less accurate than text terms ([Harrison 1997](#); [Haig and Dozier 2003b](#)), this is not always the case ([Jenuwine and Floyd 2004](#)). Also, when the subject of the search is not the main focus of the database, indexing terms may not be appropriate ([Lawrence 2011](#)).

Keywords can be extracted manually. Spreadsheets such as Excel can be used to rank keywords by frequency ([Ugolini, Neri et al. 2010](#)), but such treatment is often time-consuming because many tasks necessary for extracting keywords are not automated. Several software packages offer these options and can reduce the time required to extract keywords ([Ananiadou, Rea et al. 2009](#); [Thomas, McNaught et al. 2011](#)). Deciding which program to use will depend on criteria such as ([Neuendorf](#) ; [Lowe 2003](#); [Yang, Akers et al. 2008](#)):

- the ability to import the gold standard articles with the information contained in the different search fields of the databases;

- the flexibility offered for text analysis (different spellings, plural forms of words, different terminology, prefixes, words with the same root, or negative forms, etc.);
- the flexibility with respect to analyzing indexing terms, which are often regarded as phrases;
- the statistical tools available (frequencies, logistic regression, co-occurrence matrices, discriminant analysis, etc.).

Keywords are often selected based on their frequency, which can be calculated based on word count or record occurrence ([Jenkins 2004](#)). They should also be selected according to their relevance to the content domain for which the search filters are built.

Evaluate and combine the individual search terms

The terms selected through frequency analysis are evaluated individually in the gold standard using the indicators selected for evaluating the search filters. The best performing individual terms are then combined to form the search filters.

Search terms can be combined using Boolean operators (and, or, not) in several ways depending on the number of individual terms retained.

When the number of identified terms allows, all possible combinations can be tested using a trial-and-error model ([Harrison 1997](#)), or guided by a conceptual model of taxonomic classification ([Duff 1996](#)), or according to the conceptual scheme used in formulating the research question ([Harter 1986](#)).

When the number of terms makes assessing all possible combinations virtually impossible, other statistical approaches can be used, such as logistic regression or discriminant analysis ([White, Glanville et al. 2001](#); [Shaikh, Badgett et al. 2011](#)).

The combination of key words should avoid pitfalls known to reduce filters' efficiency, such as logic errors or improper grouping of indexing terms ([Walker, McKibbin et al. 1991](#); [Mitchell, Johnson et al. 1992](#))

Combinations of search terms are evaluated to build search filters that will suit different audiences: those who favor a very high sensitivity, those who prefer high precision, and those seeking a balance between sensitivity and precision ([Boynton, Glanville et al. 1998](#)).

Validation of the search filters

Criterion-related validity concerns the relationship or correlation between a measure and some criterion variable of interest ([Carmines and Woods 2005b](#)). It is an appropriate framework for discussing the validity of search filters, which could be described as measuring instruments. They are designed to measure the extent to which a paper indexed in a particular database could fulfill the information needs of a particular user through a quantitative scale (mostly dichotomous) used to score papers indexed in a bibliographic database ([Manning, Raghavan et al. 2008](#)). The scores are assigned in a way that distinguishes papers pertinent to a user from those that are not. Unlike reliability, which is the extent to which an empirical measurement yields consistent results in repeated trials ([Carmines and Woods 2005a](#)), validity is not a general attribute of measurement instruments, but rather, is specific to a particular purpose. It is a property of inferences, not an intrinsic property of the measuring instrument or its scores ([Messick 1989](#); [Messick 1995](#); [American Educational Research Association, American Psychological Association et al. 1999](#)). Consequently, a test could be highly valid for one purpose and not at all valid for other purposes. However, a single measuring instrument could also be a valid instrument for a range of different situations. For a single measuring instrument, there should be as many validity coefficients as there are contexts in which criterion validity is to be demonstrated ([Borneman 2010](#)).

Supporting arguments for the use of a criterion-related validity framework to assess the validity of search filters and the application of the method are presented in chapters 5 to 7. The way to establish that validity is to have an external criterion measuring the characteristics the papers are supposed to possess, to which the filter's performance could be compared ([Anastasi and Urbina 1997](#)). In that context, the filter will be the variable called predictor (**Y**) and the instrument used to decide whether a paper belongs to the gold standard will be the variable called criterion (**X**).

Consequently, the evaluative design used to assess a filter's performance is the same that will be used to assess its validity, the main difference being the indicators used. Measuring validity will essentially consist in computing the validity coefficient, which is the correlation between the predictor and the criterion. A high correlation will mean the papers selected through the predictor (the search filter) are more likely to belong to the gold standard than those not selected. The validity coefficient will establish with what degree of confidence the performance to the criterion (**X**) could be predicted from the performance to the predictor (**Y**).

When both the predictor and the criterion are measured on a binary scale, which is most often the case, their correlation is measured by the phi (ϕ) coefficient, which is a variation of Pearson's correlation coefficient (r). The phi coefficient varies between -1 and 1, the negative or positive sign indicating the direction of the relationship and the absolute value indicating the strength of the association. The phi coefficient could be interpreted as Pearson's r coefficient along the following scale: little to no correlation ($0.00 \leq |r| < 0.30$), low correlation ($0.30 \leq |r| < 0.49$), moderate correlation ($0.50 \leq |r| < 0.69$), high correlation ($0.70 \leq |r| < 0.89$), and very high correlation ($0.90 \leq |r| \leq 1.00$) ([Whitford 2005](#)).

However, the phi coefficient should be interpreted with caution, because contrary to the r coefficient, the phi coefficient rarely reaches the extreme values of ± 1 ([Whitford 2005](#)). Indeed, the phi coefficient is extremely sensitive to the difference between the mean

values of the predictor and the criterion. The greater that difference, the more the possibilities of achieving a value of 1 or -1 are reduced ([Guilford 1965](#)).

A more objective way to interpret the phi coefficient is to correct its value for the restriction of its range. Statisticians have proposed formulas to determine the maximum value (ϕ_{\max} or φ_{\max}) and the minimum value (ϕ_{\min} or φ_{\min}) the phi coefficient could take, given the means of the predictor and the criterion. The phi coefficient is then reported to the maximum value and the indicator $\frac{\varphi}{\varphi_{\max}}$ is used to have a better idea of the strength of the correlation between the predictor and the criterion. ([Guilford 1965](#); [Peduzzi, Detre et al. 1983](#); [Davenport and El-Sanhurry 1991](#); [Warrens 2008](#)).

Table 31: Indicators to assess the validity of a search filter

Name	Description	Formulas
Phi coefficient $R\phi_{XY}$	The measure of association between the two binary variables X and Y	$\frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$
Phimax	The maximum value the phi coefficient could take	$\sqrt{\frac{(c+d)}{N} \cdot \frac{(a+c)}{N}}$ if $(a+b) \geq (a+c)$ $\sqrt{\frac{(b+d)}{N} \cdot \frac{(a+b)}{N}}$ if $(a+c) \geq (a+b)$
Phimax-adjusted phi coefficient	The phi coefficient adjusted to eliminate the restriction on the range of phi	$\frac{R\phi_{XY}}{R\phi_{\max}}$

Conclusion

There are many guides available to perform classic literature searches. The performance of the information search strategies obtained is not always evaluated but those strategies could, under certain circumstances, meet the information needs of researchers, clinicians or policy-makers. ([Hammerstrøm, Wade et al. 2010](#); [Higgins, Green et al. 2011](#); [Leenaars, Hooijmans et al. 2011](#)). However, search filters are expected to play an increasingly important role in the information society ([Hersh 2009](#)). The construction and validation of optimized search filters is a task that is still relatively long, arduous, and expensive. The guide proposed here, though not exhaustive, aims to provide those who occasionally wish to engage in such a process with an overview of the different steps and information on technology resources that could help shorten the process.

Competing interests

The authors declare that they have no competing interests.

Contributors

Affaud Anaïs Tanon, François Champagne, and André-Pierre Contandriopoulos all contributed to the synthesis of the ideas. Affaud Anaïs Tanon was responsible for organizing the ideas, conducting the additional literature searches, and drafting the paper. All authors were involved in critically reviewing the article and gave final approval of the version submitted for publication.

Acknowledgments

This research was funded by the Canadian Institutes of Health Research (CIHR).

Scholarships were provided by: the AnÉIS program (Analyse et Évaluation des Interventions en Santé / Analysis and Evaluation of Health Interventions); SSHRC (Social Sciences and Humanities Research Council of Canada); GIRU (Groupe Interuniversitaire de Recherche sur les Urgences / Interuniversity Research Group on Emergencies); the Government of Côte d'Ivoire; the University of Montreal; and GRIS (Groupe de Recherche Interdisciplinaire en Santé / Interdisciplinary Team for Research in Health).

Chapitre 9: Conclusion

Le mouvement de la Médecine basée sur les évidences a depuis plus d'une trentaine d'années démontré l'importance des évidences pour soutenir les décisions prises dans le système de santé ([Claridge and Fabian 2005](#)). Mais l'accès à l'information et l'utilisation adéquate de l'information vont jouer un rôle encore plus grand dans le futur. Dans un rapport publié en 2003 et intitulé « Who will keep the public health healthy? Educating public health professionals for the 21st century », l'IOM (Institute of Medicine américain) fait un important plaidoyer pour que les compétences informationnelles soient incluses de façon plus systématique et plus approfondies dans le curriculum des professionnels de la Santé publique du 21^{ème} :

« The committee reaffirms the importance of the traditional core public health areas of epidemiology, biostatistics, environmental health, health services administration, and social and behavioral sciences. However, the committee believes that public health professionals will be better prepared to address the major health problems and challenges facing society if they achieve competency in the following eight content areas: **informatics**, genomics, communication, cultural competence, community-based participatory research, global health, policy and law, and public health ethics.» ([Gebbie, Rosenstock et al. 2003](#)).

Le premier de ces domaines, l'informatique de la santé publique ou en anglais "public health informatics", peut être défini comme l'application systématique des sciences de l'information, des sciences informatiques et de la technologie à la pratique et à l'apprentissage de la santé publique (([Yasnoff, O'Carroll et al. 2000](#)). Son rôle prédominant s'explique par le fait que l'information joue un rôle important dans le soutien de fonctions clés de la Santé Publique telles que l'appréciation des situations (qui inclue la collecte, l'analyse, l'interprétation et la communication de l'information) ou encore l'élaboration de politiques (qui repose sur une information fiable et à jour, et sur la capacité de manipuler et présenter cette information d'une manière qui fasse du sens pour les preneurs de décision) ([Gebbie, Rosenstock et al. 2003](#)).

Le repérage de l'information joue un rôle particulièrement important car les praticiens de la santé publique du futur devront mieux comprendre et mieux utiliser les systèmes d'information qui fournissent les données sur lesquelles sont basées la pratique et la recherche ([Gebbie, Rosenstock et al. 2003](#)).

Le champ du repérage de l'information est vieux de déjà plusieurs décennies ([Järvelin 2003](#)). Mais la vitesse à laquelle les technologies de l'information se développent rend nécessaire l'actualisation des approches et des méthodes. A l'heure actuelle, d'importants travaux se font pour prendre avantage des fonctionnalités offertes par le Web et développer des applications nouvelles qui minimise le degré d'expertise requis des chercheurs d'information ([Järvelin and Ingwersen 2009](#); [Järvelin 2011](#)). Mais les méthodes traditionnelles telles que celles qui consistent à développer des filtres de recherche pour les bases de données traditionnelles restent d'actualité, car les bases de données traditionnelles restent une ressource importante et l'introduction de nouvelles fonctionnalités offertes par le Web bien souvent n'affecte pas la structure fondamentale de la base de données ([FitzSimmons and Gross 2008](#)). En outre les filtres de recherche optimisés actuellement disponibles dans la littérature ne couvrent pas tous les besoins en information possibles et ne sont pas disponibles pour certains champs importants de la santé tels que la sécurité des patients.

Retour sur la problématique de départ :

L'information dans le domaine de la Santé Publique est particulièrement difficile à repérer car elle est extrêmement abondante et parce qu'elle se trouve dispersée à travers des canaux de publication, multiples et largement dispersés, en raison du caractère multidisciplinaire du champ ([Osiobe 1985](#); [Harley 2001](#); [Hood and Wilson 2001](#); [Druss and Marcus 2005](#); [Glasziou 2006](#); [Albitz 2007](#); [U.S. National Library of Medicine - National Institutes of Health 2011](#)). Les bases de données bibliographiques sont l'un des canaux les plus utilisés pour avoir recours aux évidences scientifiques dans ce domaine ([Haig and](#)

[Dozier 2003b](#)), et les acteurs de la Santé Publique sont de plus en plus encouragés à se baser sur de telles évidences pour informer leurs prises de décisions ([Brownson, Gurney et al. 1999](#); [Kohatsu, Robinson et al. 2004](#); [Anderson, Brownson et al. 2005](#); [Muir Gray 2008](#); [Brownson, Chriqui et al. 2009](#); [Brownson, Fielding et al. 2009](#); [Brownson 2011](#)).

Le recours à des filtres de recherche bibliographiques, construits en prenant avantage des connaissances produites dans les domaines des sciences de l'information et de l'informatique et adaptés aux caractéristiques des bases de données dans lesquelles l'information est répertoriée, peut contribuer de plusieurs façons à soutenir la prise de décision basée sur les évidences dans le domaine de la Santé Publique.

Tout d'abord, ils offrent un moyen effectif et systématique de repérer la littérature pertinente pour un sujet donné indexé dans une base de données bibliographiques. Leur performance peut être maximisée pour augmenter le nombre de documents pertinents, pour réduire le nombre de documents non pertinents, ou pour atteindre un certain équilibre entre le nombre de documents pertinents et le nombre de documents non pertinents qu'une base de données bibliographiques présente à un utilisateur dans le cadre de sa recherche d'information ([Boynton, Glanville et al. 1998](#)).

Les évidences produites lors de la validation des filtres de recherche bibliographiques ont pour but de fournir des réponses à des questions importantes qui accompagnent l'utilisation des filtres de recherche : quel genre de documents les filtres de recherche sont-ils censés repérer?, jusqu'à quel point peut-on affirmer qu'ils remplissent cette fonction?, jusqu'à quel point sont-ils utiles comparés par exemple à une sélection aléatoire ou à d'autres filtres disponibles pour les mêmes fins?, jusqu'à quel point couvrent-ils de façon adéquate un champ particulier de la santé publique dans la base de données bibliographique considérée?, etc. Le processus d'évaluation est donc important non seulement pour guider l'utilisation des filtres de recherche bibliographiques, mais aussi pour guider l'élaboration de filtres de recherche plus adaptés à des besoins d'information particuliers ou à la couverture d'un domaine de connaissances précis.

Ensuite, les filtres de recherche bibliographiques optimisés peuvent fournir de façon très efficiente la matière première nécessaire aux analyses quantitatives effectuées sur de grandes quantités de texte qui aident à mettre en évidence des relations et des associations qui autrement ne seraient pas décelables ([He 1999a](#); [Weber, Kors et al. 2005](#)) et produire de nouvelles évidences sur lesquelles les preneurs de décisions, les chercheurs et les cliniciens peuvent s'appuyer.

Ils peuvent être utilisés pour recueillir les données nécessaires à l'exploration de l'évolution et des frontières d'un champ particulier de connaissances, surtout lorsqu'il est fondamentalement interdisciplinaire. A titre d'exemple, Neff et Corley ([Neff and Corley 2009](#)), ont effectué une analyse textuelle quantitative basée sur près 160 000 articles pour mettre en évidence les tendances dans les méthodes utilisées et les thèmes abordés dans le champ de l'écologie sur une période de 35 ans allant de 1960 à 2005. Leur analyse leur a permis d'explorer de façon originale les relations entre politiques publiques, changements technologiques et évolution des priorités de recherche dans le domaine de l'écologie. Ils ont soulignés parmi les limites de leur étude, le fait que les articles inclus dans leur analyse provenaient d'une liste de 112 journaux catalogués dans l'édition 2005 du Web of Science Edition Journal Citation Reports comme faisant partie du champ de l'écologie. Ils ont de ce fait d'après eux laissés de côté les articles du domaine de l'écologie publiés dans les journaux plus interdisciplinaire ([Neff and Corley 2009](#)). Le fait de mieux connaître les frontières d'un champ de connaissance, les concepts qui y sont utilisés et la façon dont ils sont reliés, ainsi que les sous disciplines qui sont mobilisées, contribue à son tour à éclairer les stratégies visant à repérer les informations pertinentes dans ce domaine qui pourraient éclairer la prise de décisions.

Les filtres de recherche bibliographique peuvent être utilisés pour recueillir le matériel nécessaire à des études scientométriques visant à analyser ou à prédire la structure d'un champ particulier de connaissances pour identifier les acteurs majeurs (pays, institutions, individus) qui œuvrent dans le champ ou qui sont spécialisés autour de thèmes particuliers ([Morel, Serruya et al. 2009](#); [Kurosawa and Takama 2012](#)). L'information sur

les réseaux d'experts peut s'avérer un outil puissant pour repérer la littérature grise qui n'est pas toujours répertoriée dans les bases de données traditionnelles. En effet, les experts sont souvent consultés dans le cadre d'une recherche d'information et lorsqu'ils sont choisis de façon à refléter l'entièreté du spectre des points de vues et des compétences, leur implication maximise à la fois les chances qu'a une revue de couvrir toute la littérature pertinente mais contribue également à augmenter celles de mettre à jour les connaissances et de bénéficier de pistes sur la littérature en voie d'être publiée ([Grayson and Gomersall 2003](#)).

Ils peuvent encore être utilisés pour collecter les articles nécessaires à des études textuelles quantitatives visant à guider les politiques et les interventions autour d'une maladie spécifique. A titre d'exemple, Onyancha et Ocholla ont effectué une analyse textuelle quantitative pour identifier les caractéristiques uniques du VIH et du Sida en Afrique afin de repérer les facteurs de risques et les infections opportunistes les plus souvent associées au Sida dans le contexte africain et orienter les interventions adéquates à mettre en œuvre ([Onyancha and Ocholla 2009](#)).

D'autres applications sont possibles et ne peuvent toutes être citées, mais les deux optiques dans lesquelles les filtres de recherche bibliographiques peuvent soutenir la prise de décisions basée sur les évidences ont guidé le travail présenté dans cette thèse.

Dans le champ de la Sécurité des Patients qui illustre bien les difficultés qu'il y a à repérer l'information dans le domaine de la Santé Publique, nous nous sommes proposés d'élaborer des filtres de recherche bibliographiques qui nous permettraient de construire une importante collection, non biaisée par une définition préliminaire de la sécurité des patients, afin d'entreprendre dans un deuxième temps une analyse textuelle quantitative de cette collection d'articles qui nous permettrait d'une part d'analyser le concept de la Sécurité des Patients et d'autre part d'élaborer des filtres de recherche bibliographiques mieux adaptés aux différentes facettes du concept. Nous aurions pu nous servir de collections d'articles déjà disponibles comme celle de la National Patient Safety Foundation dont il est fait mention dans cette thèse, ou encore en construire une à partir des

suggestions de lecture disponible sur la page web du PSNet (Patient Safety Network) de l'AHRQ (Agency for HealthCare research and Quality)⁹, mais en construisant une collection directement à partir des bases de données bibliographiques, notre hypothèse est qu'il sera plus facile d'avoir une vision internationale et non biaisée du concept.

L'élaboration des filtres de recherche bibliographique s'est faite dans un premier temps en ciblant de façon délibérée les articles qui portent sur la sécurité des patients dans les organisations de santé qui ont de façon explicite été définis comme une contribution au champ de la sécurité des patients par leurs auteurs. Les filtres de recherche bibliographiques ont été élaborés pour trois types d'utilisateurs : ceux qui désirent maximiser la sensibilité, ceux qui désirent maximiser la précision et ceux qui désirent un équilibre entre sensibilité et précision.

En regard de notre désir de pouvoir détecter des changements épistémologiques dans le champ de la sécurité des patients à travers la littérature publiée, nous appartenons au premier groupe d'utilisateurs et pour cette raison, les filtres de recherche les plus intéressants à la suite de cette première phase du projet sont :

- Pour Medline: les filtres S2Med, S3Med, SP1Med et SP3Med. Le filtre S1Med nous semble moins intéressant car l'effort supplémentaire en termes d'articles non pertinents à lire (12 dans l'échantillon 1 et 28 dans l'échantillon 2 ne compense pas le gain en sensibilité de 1,20% par rapport aux filtres S2med et S3Med. Les filtres SP1Med et SP2Med ont été également considérés à ce stade parce qu'ils ont une sensibilité presque aussi importante que les filtres S2Med et S3Med dans les échantillons 1 et 2 avec une précision plus grande (voir Table 4 page 48).
- Pour Embase : les filtres S1Emb, S2Emb et S3Emb (voir Table 5 page 49).
- Pour CINAHL : les filtres S1Cin, S2Cin et S3Cin (voir Table 6 page 50).

⁹ AHRQ (Agency for Healthcare Research and Quality). "PsNet Patient safety Network - The collection: Annotated links to patient safety literature, news, and other resources." <http://psnet.ahrq.gov/collection.aspx>. (Accessed on November 2012)

L'effort de validation visait à identifier ceux des filtres de recherche qui, tout en répondant aux critères de sélection initialement choisis, permettaient de repérer des articles qui correspondent à une définition plus large de la sécurité des patients, de façon à pouvoir construire une collection non biaisée d'articles portant sur la sécurité des patients.

À l'issue de l'évaluation de la validité de critère des filtres proposés, tous les dix filtres considérés plus haut exhibaient des coefficients de validité, ajustés à leurs valeurs maximales, très élevés : les coefficients φ_{xy}/PHI_{MAX} étaient supérieurs à 0,90 dans les échantillons 1 et 2 pour tous les filtres à l'exception du filtre S3Cin dont la valeur du coefficient φ_{xy}/PHI_{MAX} était de 0,95 dans l'échantillon 1 mais de 0,85 dans l'échantillon 2 (voir Table 16 page 91). Les 10 filtres considérés plus haut exhibent une très bonne validité de critère et peuvent être considérés de façon valide comme permettant de repérer des articles répondant à nos critères de sélection.

L'évaluation de l'utilité des filtres considérés montre que les 10 filtres choisis sont fort utiles pour repérer des articles répondant à nos critères de sélection dans les deux échantillons avec des statistiques de RIOC excédant 90% dans les deux échantillons à l'exception de SP2Med (dont le RIOC est de 88% dans l'échantillon 1 et de 93% dans l'échantillon 2) et de S3Cin (dont le RIOC est de 95% dans l'échantillon 1 et de 85% dans l'échantillon 2) (voir Tableau 24 page 114). Néanmoins les différences entre les valeurs du RIOC dans les deux échantillons n'étant pas significatives, on peut dire que les 10 filtres considérés améliorent la prise de bonnes décisions de plus de 90% par rapport à une sélection aléatoire d'articles répondant à nos critères.

Finalement en testant les filtres de recherche dans le gold standard NPSF (Échantillon 3), on peut dire ceux des dix filtres préalablement sélectionnés qui offrent le plus de potentiel de repérer des articles qui correspondent à une définition plus large de la sécurité des patients (et qui par ailleurs minimisent l'effort à fournir avec le nombre d'articles non pertinents à lire) sont:

- Pour Medline : le filtre S2Med (voir Tableau 25 page 116), cependant S3Med pourrait être aussi considéré car ce filtre repère 100% des articles indexés avec le MesH « Patient Safety » (voir Figure 5 page 81 et Annexe 6 page xvii).
- Pour Embase: le filtre S2Emb (voir Tableau 26 page 118).
- Pour CINAHL: le filtre S2Cin (voir Tableau 27 page 120).

Nous comptons donc examiner deux combinaisons de trois filtres de recherche (S2Med, S2Emb et S2Cin versus S3Med, S2Emb et S2Cin) pour construire une nouvelle collection d'articles portant sur la Sécurité des Patients, non biaisée par une définition préliminaire du concept, pour des analyses textuelles quantitatives subséquentes.

Les utilisateurs potentiels des filtres de recherche proposés dans ce document pourraient utiliser les mêmes comparaisons pour choisir ceux des filtres qui répondent le mieux à leurs besoins.

Contributions de la thèse :

Les filtres de recherche bibliographique optimisés proposés dans le cadre de cette thèse sont les premiers du genre à offrir une certaine couverture des articles publiés dans le champ de la sécurité des patients. De façon délibérée, les articles que les filtres ambitionnent de repérer sont ceux qui portent sur la sécurité des patients dans les organisations de santé qui ont de façon explicite été définis comme une contribution au champ de la sécurité des patients par leurs auteurs. Ce besoin répond à notre désir de pouvoir détecter des changements épistémologiques dans le champ de la sécurité des patients à travers la littérature publiée dans le champ. Il est en effet désormais possible d'examiner à l'aide de logiciels spécialisés des milliers d'articles publiés dans un champ particulier et suivre l'évolution des concepts qui le caractérise ([He 1999a](#); [Weber, Kors et al. 2005](#))

Mais, il est très possible que les personnes qui utilisent nos filtres aient d'autres ambitions, notamment celles de repérer les articles publiés dans le champ tel qu'il est défini actuellement par les experts du champ ([Emanuel, Berwick et al. 2008](#)). Pour ces utilisateurs potentiels, il nous a semblé important de présenter la mesure dans laquelle les filtres développés pourraient être utilisés dans un tel contexte. Quoiqu'ils ne couvrent pas tout le champ de la sécurité des patients tel qu'il pourrait être défini par les suggestions de lecture offertes par exemple par la National Patient Safety Foundation américaine, les filtres de recherche bibliographique proposés, peuvent s'avérer être de précieuses aides pour initier une recherche d'information, surtout si d'autres filtres plus adaptés pour une telle couverture ne sont pas disponibles.

Cependant, même pour ceux qui seraient intéressés à repérer le même type de ressources que celles pour lesquelles nos filtres ont été développés, il est important de noter que quoique très performants, ces filtres ne doivent pas être utilisés de façon exclusive surtout si l'objectif est de faire une revue systématique.

Une autre contribution de cette thèse réside dans la réflexion sur l'évaluation de la validité de critère des filtres proposés. Le devis couramment mis en place pour construire les filtres de recherche et évaluer leur performance ([Jenkins 2004](#); [Glanville, Bayliss et al. 2008](#)) se prête bien à l'évaluation de la validité de critère avec la présence d'au moins un prédicteur (le filtre de recherche) et d'un critère (représenté par l'instrument servant à sélectionner les éléments du gold standard) et la pratique de scinder le gold standard et l'échantillon d'articles correspondant en deux groupes. Les méthodes utilisées de façon courante pour évaluer la validité de critère des instruments de mesure peuvent être adoptées mais doivent être adaptées aux caractéristiques spécifiques des filtres de recherche en tant qu'instrument de mesure, des buts dans lesquels ces filtres ont été construits et des caractéristiques des bases de données pour lesquels ils ont été construits .

Enfin, une troisième contribution réside dans un effort de vulgarisation : il nous a semblé important de résumer notre expérience sous forme de guide pour qu'elle serve

éventuellement à d'autres personnes de bâtir des guides qui répondent de façon plus spécifiques à leurs besoins.

Forces et faiblesses :

Les forces et les faiblesses de cette étude ont été abordées tout le long des précédents chapitres.

En résumé, on pourrait dire que la principale force de cette recherche réside dans le choix d'une approche de troisième génération ([Jenkins 2004](#)) pour la construction des filtres de recherche, qui permet d'introduire une certaine dose d'objectivité aux différentes étapes du processus. De façon délibérée, aucune restriction n'a été ajoutée aux filtres de recherche, de façon à laisser aux utilisateurs potentiels l'opportunité d'y ajouter les limites qui leur semblent les plus pertinentes en fonction de leurs besoins, ou de les combiner à leur guise à d'autres filtres de recherche méthodologiques. En outre, le fait de ne pas appliquer de restrictions nous permettait de faire des comparaisons avec les filtres élaborés par Westwood et al développés autour du thème de la Sécurité des Patients pour plusieurs bases de données ([Westwood, Rodgers et al. 2002](#)). Il est toutefois important de noter que parce qu'ils sont présentés sans restriction, les documents repérés à l'aide de ces filtres peuvent comprendre des articles portant de façon exclusive sur des non humains (animaux, végétaux, matériaux, etc.). Un deuxième point fort réside dans l'effort de validation des filtres proposés, et enfin un troisième point fort réside dans le choix de métriques qui soient bien adaptées aux tableaux de contingences et aux faibles valeurs moyennes des variables étudiées.

Les faiblesses de l'étude sont elles aussi liées à la méthodologie. Pour ce qui est de la construction des filtres de recherche, on peut citer le fait que l'instrument pour la sélection des articles du gold standard n'ait pas été formellement évalué pour sa fiabilité et sa validité de contenu, le fait que l'on ait pas déterminé la taille idéale pour l'échantillon et la façon de scinder cet échantillon en deux groupes qui soutiennent de façon adéquate non

seulement le développement et l'évaluation de la performance, mais aussi la validité des filtres, la méthode de sélection des mots clés individuels, uniquement basée sur la fréquence et moins sur la validité des termes qui aurait pu être bonifiée par l'utilisation de méthodes plus avancées d'analyse quantitative textuelle, la méthode de sélection des combinaisons définitives de mots clés qui devrait elle aussi tenir compte des questions de validité.

L'utilisation d'un design conceptuel pour classer les termes de recherche couplée à la décision arbitraire d'utiliser la fonction « explode » pour certains des Mesh sélectionnés a conduit à des redondances pour certains des filtres de recherche.

Dans Medline, la requête:

S1Med: (exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit/)

pourrait être simplifiée en retirant le Mesh Medical Audit/, ce qui donnerait :

S1Med: (exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/ OR Safety Management/)

En effet, les articles indexés sous ce terme sont tous repérés lorsque l'on utilise la fonction Explode (Exp) avec le Mesh « Quality of Health Care/ » en raison du fait que le terme « Medical Audit » est situé en dessous de « Quality of Health Care/ » dans l'arborescence des Mesh de Medline.

De la même façon, la requête

S2Med: (safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.) AND (exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit/)

pourrait être simplifiée en les expressions Safety Management/ et Medical Audit/ qui sont sous descripteurs respectifs de Risk Management/ et Quality of Health Care/ qui sont utilisés explosés, ce qui donnerait :

S2Med: (safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.) AND (exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/)

La requête:

SP1Med : (safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.) AND (exp *Risk Management/ OR exp *Medical Errors/ OR *Safety Management/)

pourrait être simplifiée en remplaçant l'expression (exp *Risk Management/ OR exp *Medical Errors/ OR *Safety Management/) par l'expression (exp *Risk Management/ OR exp *Medical Errors/), ce qui donnerait :

SP1Med : (safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.) AND (exp *Risk Management/ OR exp *Medical Errors/)

car le terme "Safety management" est situé en dessous du terme "Risk Management".

Dans Embase, le descripteur " Health Care System/" est un sous descripteur de « Health Care Organization/ » qui peut donc être retiré des requêtes S2Emb, S3Emb, P1Emb et P2EMB. Les requêtes simplifiées seraient donc :

S2Emb : (exp Safety/ or Safe\$.ti,ab. or Err\$.ti,ab. or Adverse.ti,ab.) AND (exp Health Care Quality/ or ((exp Safety/ or Err\$.ti,ab.) and (exp Health Care Organization/ or exp Health Care/ or Medic\$.ti,ab. or Hospital\$.ti,ab.)))

S3Emb: exp Medical Error/ or Patient Safety/ or exp Iatrogenic disease/ or ((exp Safety/ or Safe\$.ti,ab. or Err\$.ti,ab. or Adverse.ti,ab.) and (exp Health Care Quality/ or (exp *Safety/ and (exp Health Care Organization/ or exp Health Care/ or Medic\$.ti,ab. or Hospital\$.ti,ab.))))

P1Emb: (exp Safety/ or Err\$.ti,ab.) and patient\$.ti,ab. and (exp Health Care Organization/ or exp Health Care/ or Medic\$.ti,ab. or Hospital\$.ti,ab.)

P2Emb: (exp Safety/ or Err\$.ti,ab.) and (*Patient Care/ or (patient\$.ti,ab. and (exp Health Care Organization/ or exp Health Care/ or Medic\$.ti,ab. or Hospital\$.ti,ab.)))

Ces redondances n'ont pas de conséquences sur la performance des filtres proposés mais il est important de souligner que l'utilisation de la fonction «explode» de même que les redondances dans l'utilisation des Mesh peuvent avoir un impact important sur la performance des filtres de recherches ([Walker, McKibbon et al. 1991](#); [Mitchell, Johnson et al. 1992](#); [Sampson and McGowan 2006](#)). Ce type d'erreurs doit donc être pris en compte dans l'élaboration des filtres de recherche, surtout dans les bases de données qui ont un langage d'indexation organisé sous forme d'arborescence.

Pour ce qui est de la validation des filtres, il aurait été intéressant de disposer des données complètes sur l'échantillon d'articles dont est issu le gold standard NPSF de façon à procéder à une évaluation complète de la validité de critère avec ce groupe d'articles. Il aurait aussi été intéressant d'examiner plus en détail les raisons pour lesquelles certains articles du gold standard NPSF n'ont pas été repérés.

En raison des changements dans les langages d'indexation des bases de données, des instructions données aux auteurs pour la mise en forme de leur article en vue de la publication, ou même de l'évolution du langage utilisé dans un domaine particulier de la science, il est important de mettre à jour les filtres de recherche de façon assez régulière. Dans le cas précis de ce projet, la mise à jour des filtres fait partie des objectifs d'une deuxième étude visant à analyser avec plus de finesse la littérature scientifique publiée dans le champ de la Sécurité des Patients afin d'identifier les nuances autour du concept de Sécurité des Patients et les relations de ce concept avec celui de Qualité des Soins. Elle va permettre de construire des filtres plus ciblés une fois que le champ de la sécurité des patients aura été correctement délimité à l'aide d'une analyse quantitative de la littérature. Les différents concepts formant le champ pourront ainsi être positionnés les uns par rapport aux autres, et des filtres plus spécifiques pourront aussi être proposés.

Pistes de recherche :

Dans un effort pour améliorer les méthodes de construction, d'évaluation et de validation des filtres de recherche, nous pensons que les thèmes suivants devraient être examinés plus en détail :

- les méthodes d'échantillonnage qui supportent à la fois les besoins de construction et d'évaluation de la performance des filtres de recherche et les besoins liés à la validation des filtres de recherche,
- un cadre d'évaluation de la validité des filtres de recherche qui intègre de façon harmonieuse leurs caractéristiques en tant qu'instruments de mesure, les objectifs pour lesquels ils ont été construits et les caractéristiques des bases de données bibliographiques pour lesquels ils sont construits;
- Les questions de pertinence et la façon dont les critères de sélection des articles du gold standard peuvent être traduites dans le choix des mots clés composant les filtres de recherche.

Pour le champ de la sécurité des patients, de façon plus spécifique, nous pensons qu'il serait intéressant d'appliquer des méthodes plus objectives à la délimitation des frontières du champ.

Bibliographie

- Agresti, A. and B. A. Coull (1998). "Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions." The American Statistician **52**(2): 119-126.
- Albitz, R. (2007). "The what and who of information literacy and critical thinking in higher education." Portal : Libraries and the Academy **7**(1): 97-109.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education and Joint Committee on Standards for Educational and Psychological Testing (U.S.) (1999). Standards for educational and psychological testing. Washington, DC, American Educational Research Association.
- Ananiadou, S., B. Rea, N. Okazaki, R. Procter and J. Thomas (2009). "Supporting Systematic Reviews Using Text Mining." Social science computer review **27**(4): 509.
- Anastasi, A. and S. Urbina (1997). Psychological testing. Upper Saddle River, N.J., Prentice Hall.
- Andermann, A., L. Ginsburg, P. Norton, N. Arora and D. Bates (2011). "Core competencies for patient safety research: a cornerstone for global capacity strengthening." BMJ quality & safety **20**(1): 96-101.
- Anderson, L., R. Brownson, M. Fullilove, S. Teutsch, L. Novick, J. Fielding and G. Land (2005). "Evidence-based public health policy and practice: promises and limits." American Journal of Preventive Medicine **28**(5 Suppl): 226-230.
- Angoff, W. H. (1988). Validity: an evolving concept. Test validity. H. Wainer and H. I. Braun. Hillsdale, N.J., L. Erlbaum Associates: xx, 267 p.
- Aslib. Cranfield Research Project. and C. W. Cleverdon (1962). Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. Cranfield Eng.
- Avenell, A., H. H. Handoll and A. M. Grant (2001). "Lessons for search strategies from a systematic review, in The Cochrane Library, of nutritional supplementation trials in patients after hip fracture." Am J Clin Nutr **73**(3): 505-510.
- Bachmann, L. M., R. Coray, P. Estermann and G. Ter Riet (2002). "Identifying diagnostic studies in MEDLINE: reducing the number needed to read." J Am Med Inform Assoc **9**(6): 653-658.
- Baier, L. A., N. L. Wilczynski and R. B. Haynes (2010). "Tackling the growth of the obesity literature: obesity evidence spreads across many journals." International Journal of Obesity **34**(10): 1526-1530.

- Bak, G., M. Mierzwinski-Urban, H. Fitzsimmons, A. Morrison and M. Maden-Jenkins (2009). "A pragmatic critical appraisal instrument for search filters: introducing the CADTH CAI." *Health Info. Libr. J.* **26**(3): 211-219.
- Baker, G. R., P. G. Norton, V. Flintoft, R. Blais, A. Brown, J. Cox, E. Etchells, W. A. Ghali, P. Hebert, S. R. Majumdar, M. O'Beirne, L. Palacios-Derflinger, R. J. Reid, S. Sheps and R. Tamblyn (2004). "The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada." *Cmaj* **170**(11): 1678-1686.
- Bates, D. W. (2008). "Mountains in the clouds: patient safety research." *Qual Saf Health Care* **17**(3): 156-157.
- Bates, M. J. (1989). "The design of browsing and berrypicking techniques for the online search interface." *On-line review* **13**(5): 407.
- Bawden, D. and L. Robinson (2009). "The dark side of information: overload, anxiety and other paradoxes and pathologies." *Journal of Information Science* **35**(2): 180-191.
- Bax, L., L.-M. Yu, N. Ikeda and K. Moons (2007). "A systematic comparison of software dedicated to meta-analysis of causal studies." *BMC Medical Research Methodology* **7**(1): 40.
- Belkin, N. J. (1980). "Anomalous states of knowledge as a basis for information retrieval." *Revue canadienne des sciences de l'information et de bibliothéconomie - Canadian journal of information and library science* **5**: 133.
- Belkin, N. J., R. Oddy and H. Brooks (1982). "ASK for information retrieval: part 1. Background and theory." *Journal of Documentation* **38**(2).
- Benzies, K. M., S. Premji, K. A. Hayden and K. Serrett (2006). "State-of-the-evidence reviews: advantages and challenges of including grey literature." *Worldviews Evid Based Nurs* **3**(2): 55-61.
- Berwick, D. (2008). "The science of improvement." *JAMA (Chicago, Ill.)* **299**(10): 1182-1184.
- Berwick, D. M. (2005). "Broadening the view of evidence-based medicine." *Quality & Safety in Health Care* **14**(5): 315-316.
- Betn, A., L. Say, A. M. Glmezoglu, T. Allen and L. Hampson (2005). "Effectiveness of different databases in identifying studies for systematic reviews: experience from the WHO systematic review of maternal morbidity and mortality." *BMC Medical research methodology* **5**(6).
- Beverley, C. A., A. Booth and P. A. Bath (2003). "The role of the information specialist in the systematic review process: a health information case study." *Health information and libraries journal* **20**(2): 65-74.

- Bickley, S. R. and J. E. Harrison (2003). "How to.... find the evidence." *J Orthod* **30**(1): 72-78.
- Boluyt, N., L. Tjosvold, C. Lefebvre, T. P. Klassen and M. Offringa (2008). "Usefulness of systematic review search strategies in finding child health systematic reviews in MEDLINE." *Archives of Pediatrics & Adolescent Medicine* **162**(2): 111-116.
- Borneman, M. J. (2010). Criterion Validity. *Encyclopedia of Research Design*. N. J. Salkind. Thousand Oaks, USA, 'SAGE Publications': 292-297.
- Boynton, J., J. Glanville, D. McDaid and C. Lefebvre (1998). "Identifying systematic reviews in MEDLINE: developing an objective approach to search strategy design." *Journal of information science* **24**(3): 137-154.
- Bragge, J., S. Relander, A. Sunikka and P. Mannonen (2007). Enriching Literature Reviews with Computer-Assisted Research Mining. Case: Profiling Group Support Systems Research. *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*.
- Bragge, P. (2010). "Asking good clinical research questions and choosing the right study design." *Injury* **41 Suppl 1**: S3-6.
- Brennan, T. A., L. L. Leape, N. M. Laird, L. Hebert, A. R. Localio, A. G. Lawthers, J. P. Newhouse, P. C. Weiler and H. H. Hiatt (1991). "Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I." *N Engl J Med* **324**(6): 370-376.
- Brownson, R., J. Chiqui and K. Stamatakis (2009). "Understanding evidence-based public health policy." *Am J Public Health* **99**(9): 1576-1583.
- Brownson, R., J. Fielding and C. Maylahn (2009). "Evidence-based public health: a fundamental concept for public health practice." *Annual review of public health* **30**: 175-201.
- Brownson, R. C. (2011). *Evidence-based public health*. Oxford ; New York, Oxford University Press.
- Brownson, R. C., J. G. Gurney and G. H. Land (1999). "Evidence-based decision making in public health." *Journal of public health management and practice* **5**(5): 86-97.
- Buckland, M. and F. Gey (1994). "The Relationship between Recall and Precision." *Journal of the American Society for Information Science* **45**(1): 12-19.
- Buckland, M. K. (1997). "What is a "Document"?" *Journal of the American Society for Information Science* **48**(9): 804.
- Burns, P. B. and K. C. Chung (2010). "Developing good clinical questions and finding the best evidence to answer those questions." *Plastic & Reconstructive Surgery* **126**(2): 613-618.

- Cairney, J. and D. L. Streiner (2011). "Using relative improvement over chance (RIOCI) to examine agreement between tests: Three case examples using studies of developmental coordination disorder (DCD) in children." Research in Developmental Disabilities **32**(1): 87-92.
- Carmines, E. G. and J. A. Woods (2005a). Reliability Assessment. Encyclopedia of Social Measurement. K. Kempf-Leonard. New York, Elsevier: 361-365.
- Carmines, E. G. and J. A. Woods (2005b). Validity Assessment. Encyclopedia of Social Measurement. K. Kempf-Leonard. New York, Elsevier: 933-937.
- Champagne, F., A.-P. Contandriopoulos, M. P. Pomey and A. Tanon (2006). Élaboration d'un modèle conceptuel d'analyse de la sécurité dans les organisations de santé, GRIS Université de Montréal.
- Chung, K. and A. Ram (2009). "Evidence-based medicine: the fourth revolution in American medicine?" Plastic and reconstructive surgery **123**(1): 389-398.
- Claridge, J. A. and T. C. Fabian (2005). "History and development of evidence-based medicine." World J Surg **29**(5): 547-553.
- Cleverdon, C. (1964). The Testing and Evaluation of the Operating Efficiency of the Intellectual Stages of Information Retrieval Systems. Working Paper No. 4. The International Study Conference on Classification Research. Elsinore, Denmark.
- Cleverdon, C. and M. Keen (1966). "Factors Affecting the Performance of Indexing Systems." ASLIB, Cranfield Research Project **2**: 37-59.
- Cleverdon, C. W. (1972). "On the inverse relationship of recall and precision." Journal of Documentation **28**(3): 195.
- Cobus, L. (2008). "Integrating information literacy into the education of public health professionals: roles for librarians and the library." Journal of the Medical Library Association **96**(1): 28-33.
- Cohen, A., P. Z. Stavri and W. Hersh (2004). "A categorization and analysis of the criticisms of Evidence-Based Medicine." International journal of medical informatics **73**(1): 35-43.
- Copas, J. B. and R. Loeber (1990). "Relative improvement over chance (RIOCI) for 2x2 tables." British Journal of Mathematical and Statistical Psychology **43**(2): 293-307.
- Corrall, C. J., P. Wyer, L. Zick and C. Bockrath (2002). "Evidence-based emergency medicine. How to find evidence when you need it, part 1: databases, search programs, and strategies." Annals of emergency medicine **39**(3): 302-306.
- Cosijn, E. (2009). Relevance Judgments and Measurements. Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 4512-4519.
- Cowie, J. and W. Lehnert (1996). "Information extraction." Communications of the ACM **39**(1): 80-91.

- Croft, W. B. (1995) "What do people want from Information Retrieval (the Top 10 Research Issues for Companies that Use and Sell IR Systems)." D-Lib Magazine DOI: <http://www.dlib.org/dlib/november95/11croft.html>.
- Cronbach, L. J. (1988). Five perspectives on validity argument. Test validity. H. Wainer, H. I. Braun and Educational Testing Service. Hillsdale, N.J., L. Erlbaum Associates: xx, 267 p.
- Cronbach, L. J. and G. C. Gleser (1965). Psychological tests and personnel decisions. Urbana,, University of Illinois Press.
- Cronbach, L. J. and P. E. Meehl (1955). "Construct validity in psychological tests." Psychol Bull **52**(4): 281-302.
- Crumley, E. T., N. Wiebe, K. Cramer, T. P. Klassen and L. Hartling (2005). "Which resources should be used to identify RCT/CCTs for systematic reviews: a systematic review." BMC Med Res Methodol **5**: 24.
- Damarell, R., J. Tieman, R. Sladek and P. Davidson (2011). "Development of a heart failure filter for Medline: an objective approach using evidence-based clinical practice guidelines as an alternative to hand searching." BMC Medical research methodology **11**: 12-12.
- Davenport, E. C. and N. A. El-Sanhurry (1991). "Phi/Phimax: Review and Synthesis." Educational and psychological measurement **51**(4): 821-828.
- Davies, J. M., P. C. Hebert, C. Hoffman, Canadian Patient Safety Institute and Royal College of Physicians and Surgeons of Canada (2003). The Canadian patient safety dictionary. Calgary, Royal College of Physicians and Surgeons of Canada.
- Davies, K. (2007). "The information-seeking behaviour of doctors: a review of the evidence." Health Info. Libr. J. **24**(2): 78-94.
- Davis, P., R. Lay-Yee, R. Briant, W. Ali, A. Scott and S. Schug (2002). "Adverse events in New Zealand public hospitals I: occurrence and impact.[see comment]." New Zealand Medical Journal **115**(1167): U271.
- Davis, P., R. Lay-Yee, R. Briant, W. Ali, A. Scott and S. Schug (2003). "Adverse events in New Zealand public hospitals II: preventability and clinical context.[see comment]." New Zealand Medical Journal **116**(1183): U624.
- De Groot, S. L. and J. L. Dorsch (2003). "Measuring use patterns of online journals and databases." J Med Libr Assoc **91**(2): 231-240.
- Delamere, F. M. and H. C. Williams (2001). "How can hand searching the dermatological literature benefit people with skin problems?" Archives of dermatology **137**(3): 332-335.
- DeVellis, R. F. (2005). Inter-Rater Reliability. Encyclopedia of Social Measurement. K. Kempf-Leonard. New York, Elsevier: 317-322.

- Doms, A. and M. Schroeder (2005). "GoPubMed: exploring PubMed with the Gene Ontology." Nucleic acids research **33**(Web Server issue): W783-786.
- Druss, B. and S. Marcus (2005). "Growth and decentralization of the medical literature: implications for evidence-based medicine." Journal of the Medical Library Association **93**(4): 499-501.
- Duff, A. (1996). "The literature search: a library-based model for information skills instruction." Library review **45**(4): 14.
- Egger, M., P. Juni, C. Bartlett, F. Hoenstein and J. Sterne (2003). "How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study." Health Technol Assess **7**(1): 1-76.
- Elder, N. C. and S. M. Dovey (2002). "Classification of medical errors and preventable adverse events in primary care: a synthesis of the literature." Journal of Family Practice **51**(11): 927-932.
- Elsevier. (2012). "Embase Biomedical answers What is Embase?", from <http://www.embase.com/info/what-embase>.
- Emanuel, L., D. Berwick, J. Conway, J. Combes, M. Hatlie, L. Leape, J. Reason, P. Schyve, C. Vincent and M. Walton (2008). What Exactly Is Patient Safety? Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 1: Assessment). Henriksen K, Battles JB, Keyes MA and Grady ML. Rockville (MD), Agency for Healthcare Research and Quality.
- Farrington, D. P. and R. Loeber (1989). "Relative improvement over chance (RIOCI) and phi as measures of predictive efficiency and strength of association in 2x2 tables." Journal of Quantitative Criminology **5**(3): 201-213.
- Fineout-Overholt, E., S. Hofstetter, L. Shell and L. Johnston (2005). "Teaching EBP: getting to the gold: how to search for the best evidence." Worldviews on evidence-based nursing **2**(4): 207-211.
- Fisher, K. E., C. Landry and C. Naumer (2006). "Social spaces, casual interactions, meaningful exchanges: "information ground" characteristics based on the college student experience." Information research **12**(2): np.
- FitzSimmons, M. and V. A. Gross (2008). "A Review of the OvidSP Platform." Medical Reference Services Quarterly **27**(4): 394-405.
- Flynn, M. G. and C. McGuinness (2011). "Hospital clinicians' information behaviour and attitudes towards the 'Clinical Informationist': an Irish survey." Health Information & Libraries Journal **28**(1): 23-32.
- Ford, N. (1999). "Information retrieval for evidence based decision making." The Journal of documentation **55**(4): 385-401.

- Forster, A. J., H. J. Murff, J. F. Peterson, T. K. Gandhi and D. W. Bates (2003). "The incidence and severity of adverse events affecting patients after discharge from the hospital.[see comment]." Annals of Internal Medicine **138**(3): 161-167.
- Furr, R. M. and V. R. Bacharach (2008). Psychometrics : an introduction. Los Angeles, Sage Publications.
- Gardois, P., R. Calabrese, N. Colombi, A. Deplano, C. Lingua, F. Longo, M. C. Villanacci, R. Miniero and A. Piga (2011). "Effectiveness of bibliographic searches performed by paediatric residents and interns assisted by librarians. A randomised controlled trial." Health Information & Libraries Journal **28**(4): 273-284.
- Garg, A. X., A. V. Iansavichus, M. Kastner, L. A. Walters, N. Wilczynski, K. A. McKibbin, R. C. Yang, F. Rehman and R. B. Haynes (2006). "Lost in publication: Half of all renal practice evidence is published in non-renal journals." Kidney international **70**(11): 1995-2005.
- Gebbie, K. M., L. Rosenstock and L. M. Hernandez (2003). Who will keep the public healthy? : educating public health professionals for the 21st century. Washington, D.C., National Academies Press.
- Glanville, J., S. Bayliss, A. Booth, Y. Dundar, H. Fernandes, N. D. Fleeman, L. Foster, C. Fraser, A. Fry-Smith, S. Golder, C. Lefebvre, C. Miller, S. Paisley, L. Payne, A. Price, K. Welch, J. Glanville, S. Bayliss, A. Booth, Y. Dundar, H. Fernandes, N. D. Fleeman, L. Foster, C. Fraser, A. Fry-Smith, S. Golder, C. Lefebvre, C. Miller, S. Paisley, L. Payne, A. Price and K. Welch (2008). "So many filters, so little time: the development of a search filter appraisal checklist." Journal of the Medical Library Association **96**(4): 356-361.
- Glasziou, P. (2006). "Managing the evidence flood." Surg Clin North Am **86**(1): 193-199, xi.
- Golder, S., Y. Loke and H. McIntosh (2008). "Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects." Journal of Clinical Epidemiology **61**(5): 440-448.
- Golder, S., H. M. McIntosh, S. Duffy and J. Glanville (2006). "Developing efficient search strategies to identify reports of adverse effects in Medline and Embase." Health Info. Libr. J. **23**(1): 3-12.
- Goodwin, L. D. (2002). "Changing conceptions of measurement validity: an update on the new standards." Journal of Nursing Education **41**(3): 100-106.
- Grant, M. J. (2004). "How does your searching grow? A survey of search preferences and the use of optimal search strategies in the identification of qualitative research." Health Info. Libr. J. **21**(1): 21-32.

- Grayson, L. and A. Gomersall (2003). Working Paper 19 - A difficult business: finding the evidence for social science review. ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary University of London.
- Green, M. L. and T. R. Ruff (2005). "Why do residents fail to answer their clinical questions? A qualitative study of barriers to practicing evidence-based medicine." Academic medicine **80**(2): 176-182.
- Greenhalgh, T. (1997). "How to read a paper. The Medline database." BMJ **315**(7101): 180-183.
- Guilford, J. P. (1965). "The minimal Phi Coefficient and the Maximal Phi." Educational and psychological measurement **25**(1): 3-8.
- Gwyneth, T. (2003). Databases. International encyclopedia of information and library science. J. Feather and R. P. Sturges. London ; New York, Routledge: 126-127.
- Haig, A. and M. Dozier (2003a). "BEME Guide no 3: systematic searching for evidence in medical education--Part 1: Sources of information." Medical teacher **25**(4): 352-363.
- Haig, A. and M. Dozier (2003b). "BEME guide no. 3: systematic searching for evidence in medical education--part 2: constructing searches." Medical Teacher **25**(5): 463-484.
- Hall, J. G., L. Bainbridge, A. Buchan, A. Cribb, J. Drummond, C. Gyles, T. P. Hicks, C. McWilliam, B. Paterson, P. A. Ratner, E. Skarakis-Doyle and P. Solomon (2006). "A meeting of minds: interdisciplinary research in the health sciences in Canada." Canadian Medical Association Journal **175**(7): 763-771.
- Hammerstrøm, K., A. Wade and A.-M. K. Jørgensen (2010). Searching for studies: A guide to information retrieval for Campbell Systematic Reviews Campbell Systematic Reviews. Campbell Systematic Reviews, The Campbell Collaboration. **Supplement 1**: 74.
- Hammond, C. C. and S. W. Brown (2008). "Citation Searching: Search Smarter & Find More." Computers in Libraries **28**(5): 10-61.
- Hammons, T., N. F. Piland, S. D. Small, M. J. Hatlie and H. R. Burstin (2003). "Ambulatory patient safety. What we know and need to know." Journal of Ambulatory Care Management **26**(1): 63-82.
- Harley, B. (2001). "Freshmen, information literacy, critical thinking and values." Reference services review **29**(4): 301.
- Harris, M. (2005). "The librarian's roles in the systematic review process: a case study." Journal of the Medical Library Association **93**(1): 81-87.
- Harrison, J. (1997). "Designing a search strategy to identify and retrieve articles on evidence-based health care using MEDLINE." Health libraries review **14**(1): 33-42.

- Hart, C. (2001). Doing a literature search : a comprehensive guide for the social sciences. London ; Thousand Oaks, Calif., Sage.
- Harter, S. P. (1986). Online information retrieval : concepts, principles, and techniques. Orlando, Academic Press.
- Harter, S. P. and C. A. Hert (1997). "Evaluation of Information Retrieval Systems: Approaches, Issues, and Methods." Annual Review of Information Science and Technology **32**: 3-94.
- Haynes, R. B. (2006). "Forming research questions." Journal of Clinical Epidemiology **59**(9): 881-886.
- Haynes, R. B., N. Wilczynski, K. A. McKibbin, C. J. Walker and J. C. Sinclair (1994). "Developing optimal search strategies for detecting clinically sound studies in MEDLINE." Journal of the American Medical Informatics Association **1**(6): 447-458.
- Haynes, R. B. and N. L. Wilczynski (2004). "Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey." BMJ **328**(7447): 1040.
- Haynes, S. N., D. C. S. Richard and E. S. Kubany (1995). "Content validity in psychological assessment: A functional approach to concepts and methods." Psychological assessment **7**(3): 238-247.
- He, Q. (1999a). "Knowledge Discovery Through Co-Word Analysis." Library Trends **48**(1): 133.
- He, Q. (1999b). "Knowledge discovery through co-word analysis." Library Trends **48**(1): 133-159.
- Helmer, D., I. Savoie, C. Green and A. Kazanjian (2001). "Evidence-based practice: extending the search to find material for the systematic review." Bull Med Libr Assoc **89**(4): 346-352.
- Hemminger, B., D. Lu, K. Vaughan and S. Adams (2007). "Information seeking behavior of academic scientists." Journal of the American Society for Information Science and Technology **58**(14): 2205-2225.
- Hersh, W., M. K. Crabtree, D. Hickam, L. Sacherek, C. Friedman, P. Tidmarsh, C. Mosbaek and D. Kraemer (2002). "Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions." Journal of the American Medical Informatics Association **9**(3): 283-293.
- Hersh, W. R. (2009). Information retrieval : a health and biomedical perspective. New York, NY, Springer.

- Higgins, J. P. T., S. Green and Cochrane Collaboration. (2011). *Cochrane handbook for systematic reviews of interventions*. Cochrane book series. Chichester, England ; Hoboken, NJ, Wiley-Blackwell. **Version 5.1.0 [updated March 2011]**: xxi, 649 p.
- Hood, W. W. and C. S. Wilson (2001). "The scatter of documents over databases in different subject domains: How many databases are needed?" Journal of the American Society for Information Science and Technology **52**(14): 1242-1254.
- Hopewell, S., M. J. Clarke, C. Lefebvre and R. W. Scherer (2007) "Handsearching versus electronic searching to identify reports of randomized trials." Cochrane Database Syst Rev, 1469-1493X DOI: 10.1002/14651858.
- Howard, D. W. (2009). Relevance in Theory. Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 4498-4511.
- Huber, J. T., J. A. Boorkman and J. C. Blackwell (2008). Introduction to reference sources in the health sciences. New York, Neal-Schuman Publishers.
- Huehns, T. and M. Fletcher (2010). "Patient safety research in the UK and beyond: 10 years of research and policy to make health care safer." J Health Serv Res policy **15**: 87-91.
- Hunter, L. and K. B. Cohen (2006). "Biomedical language processing: what's beyond PubMed?" Molecular cell **21**(5): 589-594.
- Iansavichus, A. V., R. B. Haynes, S. Z. Shariff, M. Weir, N. L. Wilczynski, A. McKibbin, F. Rehman and A. X. Garg (2010). "Optimal search filters for renal information in EMBASE." American Journal of Kidney Diseases **56**(1): 14-22.
- Jackson, R. R., W. W. Baird, L. L. Davis-Reynolds, C. C. Smith, S. S. Blackburn and J. J. Allsebrook (2007). "The information requirements and information-seeking behaviours of health and social care professionals providing care to children with health care needs: a pilot study." Health information and libraries journal **24**(2): 95-102.
- Jansen, B. J. and S. Y. Rieh (2010). "The seventeen theoretical constructs of information searching and information retrieval." Journal of the American Society for Information Science and Technology **61**(8): 1517.
- Järvelin, K. (2003). Information retrieval. International encyclopedia of information and library science. J. Feather and R. P. Sturges. London ; New York, Routledge: 126-127.
- Järvelin, K. (2011). "IR Research: Systems, Interaction, Evaluation and Theories " Lecture Notes in Computer Science **6611**(Advances in Information Retrieval): 1-3.
- Järvelin, K. and P. Ingwersen (2009). User-Oriented and Cognitive Models of Information Retrieval. Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 5521-5534.

- Jenkins, M. (2004). "Evaluation of methodological search filters: a review." *Health Info. Libr. J.* **21**(3): 148-163.
- Jenkins, M. and F. Johnson (2004). "Awareness, use and opinions of methodological search filters used for the retrieval of evidence-based medical literature: a questionnaire survey." *Health Info Libr J* **21**(1): 33-43.
- Jenuwine, E. S. and J. A. Floyd (2004). "Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals." *Journal of the Medical Library Association* **92**(3): 349-353.
- Kagolovsky, Y. and J. R. Moehr (2003). "Current Status of the Evaluation of Information Retrieval." *Journal of Medical Systems* **27**(5): 409-424.
- Kane, M. T. (2001). "Current Concerns in Validity Theory." *Journal of Educational Measurement* **38**(4): 319-342.
- Kastner, M., N. L. Wilczynski, A. K. McKibbin, A. X. Garg and R. B. Haynes (2009). "Diagnostic test systematic reviews: bibliographic search filters ("Clinical Queries") for diagnostic accuracy studies perform well." *Journal of Clinical Epidemiology* **62**(9): 974-981.
- Kelly, L. and N. St Pierre-Hansen (2008). "So many databases, such little clarity: Searching the literature for the topic aboriginal." *Can Fam Physician* **54**(11): 1572-1573.
- Kim, J.-J. and D. Rebholz-Schuhmann (2008). "Categorization of services for seeking information in biomedical literature: a typology for improvement of practice." *Briefings in bioinformatics* **9**(6): 452-465.
- King, D. N. (1987). "The contribution of hospital library information services to clinical care: a study in eight hospitals." *Bulletin of the Medical Library Association* **75**(4): 291-301.
- Knottnerus, J. A. and F. Buntinx (2009). *The evidence base of clinical diagnosis : theory and methods of diagnostic research*. Oxford ; Hoboken, NJ, Wiley-Blackwell Pub./BMJ Books.
- Kohatsu, N. D., J. G. Robinson and J. C. Torner (2004). "Evidence-based public health: an evolving concept." *American Journal of Preventive Medicine* **27**(5): 417-421.
- Kohn, L. T., J. M. Corrigan and M. S. Donaldson (1999). *To Err is human: building a safer health system*. Washington DC, National Academy Press.
- Korfhage, R. R. (1997). *Information storage and retrieval*. New York, Wiley Computer Pub.
- Kostagiolas, P. A., V. A. Aggelopoulou and D. Niakas (2011). "A study of the information seeking behaviour of hospital pharmacists: empirical evidence from Greece." *Health Information & Libraries Journal* **28**(4): 302-312.

- Kotz, S., C. B. Read and D. L. Banks (1997). Encyclopedia of statistical sciences. Update. New York, Wiley.
- Kramer, B. M. R. (2011) "Information Retrieval: Literature Searching in Today's Information Landscape." Hypothesis Journal **8**.
- Kurosawa, T. and Y. Takama (2012). "Co-Authorship Networks Visualization System for Supporting Survey of Researchers' Future Activities." Journal of Emerging Technologies in Web Intelligence **4**(1): 3.
- Lancaster, L. and V. Gale (2003). Pertinence and Relevance. Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 2307 - 2315.
- Larson, R. R. (2009). Information Retrieval Systems. Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 2553-2563.
- Lawrence, D. (2011). "Controlled search term vocabularies for finding articles relevant to injury prevention and safety promotion." Injury Prevention **17**(4): 260-265.
- Lawrence, D. W. (2008). "What is lost when searching only one literature database for articles relevant to injury prevention and safety promotion?" Injury Prevention **14**(6): 401-404.
- Lawrence, J. C. (2007). "Techniques for searching the CINAHL database using the EBSCO interface." AORN Journal **85**(4): 779.
- Leape, L., D. Berwick and D. Bates (2002). "What practices will most improve safety? Evidence-based medicine meets patient safety." JAMA (Chicago, Ill.) **288**(4): 501-507.
- Leape, L. L. (2008). "Scope of problem and history of patient safety." Obstet Gynecol Clin North Am **35**(1): 1-10.
- Leape, L. L., T. A. Brennan, N. Laird, A. G. Lawthers, A. R. Localio, B. A. Barnes, L. Hebert, J. P. Newhouse, P. C. Weiler and H. Hiatt (1991). "The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II." N Engl J Med **324**(6): 377-384.
- Leeftang, M. M., R. J. Scholten, A. W. Rutjes, J. B. Reitsma and P. M. Bossuyt (2006). "Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies." Journal of Clinical Epidemiology **59**(3): 234-240.
- Leenaars, M., C. Hooijmans, N. van Veggel, G. Ter Riet, M. Leeftang, L. Hooft, G. J. van der Wilt, A. Tillema and M. Ritskes Hoitinga (2011). "A step-by-step guide to systematically identify all relevant animal studies." Laboratory animals.
- Ley, P. (1972). Quantitative aspects of psychological assessment: an introduction. London,, Duckworth.

- Lipowski, E. E. (2008). "Developing great research questions." American Journal of Health-System Pharmacy **65**(17): 1667-1670.
- Littell, J. H., J. Corcoran and V. K. Pillai (2008). Systematic reviews and meta-analysis. Oxford ; New York, Oxford University Press.
- Loeber, R. and T. Dishion (1983). "Early predictors of male delinquency: a review." Psychol Bull **94**(1): 68-99.
- Lowe, W. (2003). Content Analysis Software: A Review (Technical Report for the Identity Project, Weatherhead Center for International Affairs, Harvard University)
Identity Project, Weatherhead Center for International Affairs, Harvard University. .
- Lu, Z., W. Kim and W. J. Wilbur (2009). "Evaluation of Query Expansion Using MeSH in PubMed." Information Retrieval **12**(1): 69-80.
- MacDonald, J., P. Bath and A. Booth (2011). "Information overload and information poverty: challenges for healthcare services managers?" Journal of Documentation **67**(2): 238-263.
- MacDonald, R. D., B. A. Banks and M. Morrison (2008). "Epidemiology of adverse events in air medical transport." Academic Emergency Medicine **15**(10): 923-931.
- Major, M., S. Warren and C. Flores Mir (2009). "Survey of systematic review authors in dentistry: challenges in methodology and reporting." Journal of dental education **73**(4): 471-482.
- Manning, C. D., P. Raghavan and H. Schütze (2008). Introduction to information retrieval. New York, Cambridge University Press.
- Marshall, J. G. (1992). "The impact of the hospital library on clinical decision making: the Rochester study." Bulletin of the Medical Library Association **80**(2): 169-178.
- McAuley, L., B. Pham, P. Tugwell and D. Moher (2000). "Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses?" Lancet.
- McCord, J. (1980). Patterns of deviance. Human functioning in longitudinal perspective : studies of normal and psychopathic populations. S. B. Sells, R. Crandall, M. Roff, J. S. Strauss and W. Pollin. Baltimore, Williams & Wilkins: xv, 283 p.
- McDonald, M. P. (2005). Validity, Data Sources. Encyclopedia of Social Measurement. K. Kempf-Leonard. New York, Elsevier: 939-948.
- McGartland Rubio, D. (2005). Content Validity. Encyclopedia of Social Measurement. New York, Elsevier: 495-498.
- McGowan, J. and M. Sampson (2005). "Systematic reviews need systematic searchers." Journal of the Medical Library Association **93**(1): 74.

- McKibbin, K., N. Wilczynski and R. Haynes (2004). "What do evidence-based secondary journals tell us about the publication of clinically important articles in primary healthcare journals?" BMC medicine **2**: 33-33.
- McKibbin, K. A. (1998). "Evidence-based practice." Bulletin of the Medical Library Association **86**(3): 396-401.
- McKibbin, K. A., C. Lokker, N. L. Wilczynski, R. B. Haynes, D. Ciliska, M. Dobbins, D. A. Davis and S. E. Straus (2012). "Search filters can find some but not all knowledge translation articles in MEDLINE: an analytic survey." Journal of Clinical Epidemiology **65**(6): 651-659.
- McKibbin, K. A. and C. J. Walker-Dilks (1995). "The quality and impact of MEDLINE searches performed by end users." Health libraries review **12**(3): 191-200.
- Meadow, C. T. (1967). The analysis of information systems; a programmer's introduction to information retrieval. New York, Wiley.
- Melnyk, B. and E. Fineout-Overholt (2002). "Key steps in implementing evidence-based practice: asking compelling, searchable questions and searching for the best evidence." Pediatric nursing **28**(3): 262-263, 266.
- Messick, S. (1989). Validity. Educational measurement (3rd ed.). American Council on Education, American Council on Education: 13-103.
- Messick, S. (1995). "Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning." American Psychologist **50**(9): 741-749.
- Michel, P., J. L. Quenon, A. Djihoud, S. Tricaud-Vialle and A. M. de Sarasqueta (2007). "French national survey of inpatient adverse events prospectively assessed with ward staff." Quality & Safety in Health Care **16**(5): 369-377.
- Miller, F., M. Giacomini, J. Abelson, J. Hurley, A. Iannantuono, P. Smith, J. Eyles and B. O'Brien (2001). A review of key literatures for the Study of Health System Safety, Error, and Injury. A report submitted to Policy division, Health Policy & Communication Branch. Health Canada., McMaster University.
- Mitchell, J. A., E. D. Johnson, J. E. Hewett and V. K. Proud (1992). "Medical students using Grateful Med: analysis of failed searches and a six-month follow-up study." Computers and biomedical research **25**(1): 43-55.
- Montori, V. and G. Guyatt (2008). "Progress in evidence-based medicine." JAMA (Chicago, Ill.) **300**(15): 1814-1816.
- Morel, C., S. Serruya, G. Penna and R. Guimaraes (2009). "Co-authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programs on Neglected Diseases." PLoS Neglected Tropical Diseases **3**(8): e501.

- Muir Gray, J. A. (2008). Evidence-based healthcare and public health : how to practice and teach evidence-based decision making. Edinburgh ; New York, Churchill Livingstone.
- Munck, G. L. and J. Verkuilen (2005). Research Designs. Encyclopedia of Social Measurement. K. Kempf-Leonard. New York, Elsevier: 385-395.
- National Patient Safety Foundation. Current Awareness Retrieved October, 21, 2009, from <http://www.npsf.org/rc/pubs/ca/>.
- National Patient Safety Foundation resource page. "Current Awareness Literature Alert." Retrieved May 2011, 2011, from <http://www.npsf.org/rc/pubs/ca/>.
- Naumer, C. M. and K. E. Fisher (2009). Information Needs. Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 2452-2458.
- Neff, M. and E. Corley (2009). "35 years and 160,000 articles: A bibliometric exploration of the evolution of ecology." Scientometrics **80**(3): 657-682.
- Neuendorf, K. A. "Quantitative Text Analysis Programs." Retrieved January 02, 2012, 2012, from <http://academic.csuohio.edu/kneuendorf/content/cpuca/qtap.htm>.
- Neuendorf, K. A. (2002). The content analysis guidebook. Thousand Oaks, Calif., Sage Publications.
- Newby, J. (2011). "Entering Unfamiliar Territory Building an Information Literacy Course for Graduate Students in Interdisciplinary Areas." Reference & User Services Quarterly **50**(3): 224-229.
- Nunnally, J. C. and I. H. Bernstein (1994). Psychometric theory. New York, McGraw-Hill.
- O'Brien, P. (2004). The electronic systematic review handbook : practical concepts and methods for electronic screening and data extraction. Ottawa, TrialStat Corporation.
- Office Québécois de la langue française. (2000). "Le grand dictionnaire terminologique." from http://w3.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp.
- Onyanha, O. and D. Ocholla (2009). "Is HIV/AIDS in Africa distinct? What can we learn from an analysis of the literature?" Scientometrics **79**(2): 277-296.
- Osiobe, S. A. (1985). "Use of Information resources by health professionals - A review of the literature." Social Science & Medicine **21**(9): 965-973.
- Patrick, T., G. Demiris, L. Folk, D. Moxley, J. Mitchell and D. Tao (2004). "Evidence-based retrieval in evidence-based medicine." Journal of the Medical Library Association **92**(2): 196-199.
- Pedhazur, E. J. and L. P. Schmelkin (1991). Measurement, design, and analysis : an integrated approach. Hillsdale, N.J., Lawrence Erlbaum Associates.

- Peduzzi, P. N., K. M. Detre and Y. K. Chan (1983). "Upper and lower bounds for correlations in 2 x 2 tables - Revisited." Journal of Chronic Diseases **36**(7): 491-496.
- Petrosino, A. J. (1995). "Specifying inclusion criteria for a metaanalysis - Lessons and illustrations from a quantitative synthesis of crime reduction experiments." Evaluation review **19**(3): 274-293.
- Pettersen, N. and Ebrary (2000). Évaluation du potentiel humain dans les organisations élaboration et validation d'instruments de mesure. Sainte-Foy Que., Presses de l'Université du Québec.
- Pickett, K. M. (2008). "Reaching beyond MEDLINE: A Beginner's Overview of Electronic Biomedical Resources." Journal of Hospital Librarianship **8**(4): 398-410.
- Plake, C., T. Schiemann, M. Pankalla, J. Hakenberg and U. Leser (2006). "AliBaba: PubMed as a graph." Bioinformatics **22**(19): 2444-2445.
- Powers, D. (2011). "Evaluation: from precision, recall and F-measure to Roc, informedness, markedness & Correlation." Journal of Machine Learning Technologies **2**(1): 37-63.
- Raffle, A. E. and J. A. Muir Gray (2007). Screening : evidence and practice. Oxford ; New York, Oxford University Press.
- Ramsay, S. (2001). "Johns Hopkins takes responsibility for volunteer's death." The Lancet - British Edition **358**(9277): 213-213.
- Randomness and Integrity Services Limited. "Random.org : True Random Number Service." 2011, from <http://www.random.org/>.
- Ranji, S. and K. Shojania (2008). "Implementing patient safety interventions in your hospital: What to try and what to avoid." The Medical clinics of North America **92**(2): 275-+.
- Ritchie, G., J. Glanville and C. Lefebvre (2007). "Do published search filters to identify diagnostic test accuracy studies perform adequately?" Health Information & Libraries Journal **24**(3): 188-192.
- Rowley, J. E. and R. J. Hartley (2008). Organizing knowledge : an introduction to managing access to information. Aldershot, England ; Burlington, VT, Ashgate.
- Runciman, W. B., G. R. Baker, P. Michel, I. L. Jauregui, R. J. Lilford, A. Andermann, R. Flin and W. B. Weeks (2008). "The epistemology of patient safety research." Int J Evid Based Healthc **6**(4): 476-486.
- Ruthven, I. (2005). Integration Approaches to Relevance. New Directions in Cognitive Information Retrieval. A. Spink and C. Cole, Springer Netherlands. **19**: 61-80.

- Sackett, D. L., S. E. Straus, W. S. Richardson, W. Rosenberg and R. B. Haynes (2000). Evidence-based medicine : how to practice and teach EBM. Edinburgh ; New York, Churchill Livingstone.
- Salton, G. (1965). "The evaluation of automatic retrieval procedures— selected test results using the SMART system." American Documentation **16**(3): 209-222.
- Salton, G. (2009). SMART System: 1961-1976. Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 4770-4790.
- Salton, G. and M. J. McGill (1983). Introduction to modern information retrieval. New York, McGraw-Hill.
- Sampson, M. and J. McGowan (2006). "Errors in search strategies were identified by type and frequency." Journal of Clinical Epidemiology **59**(10): 1057-1063.
- Sampson, M., K. G. Shojania, C. Garritty, T. Horsley, M. Ocampo and D. Moher (2008). "Systematic reviews can be produced and published faster." J Clin Epidemiol **61**(6): 531-536.
- Sampson, M., L. Zhang, A. Morrison, N. J. Barrowman, T. J. Clifford, R. W. Platt, T. P. Klassen and D. Moher (2006) "An alternative to the hand searching gold standard: validating methodological search filters using relative recall." BMC Med Res Methodol **6**, 33 DOI: 1471-2288-6-33 [pii] 10.1186/1471-2288-6-33.
- Sandars, J. and A. Esmail (2003). "The frequency and nature of medical error in primary care: understanding the diversity across studies." Fam Pract **20**(3): 231-236.
- Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems Foundations and Trends in Information Retrieval **4**: 247-375.
- Saracevic, T. (2007). "Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance." J Am Soc Inf Sci Technol **58**(13): 1915-1933.
- Saracevic, T. (2008). "Effects of inconsistent relevance judgments on information retrieval test results: A historical perspective." Library Trends **56**(4): 763-783.
- Saranto, K. and E. Hovenga (2004). "Information literacy - what it is about? Literature review of the concept and the context." International Journal of Medical Informatics **73**(6): 503-513.
- Schlosser, R., O. Wendt, S. Bhavnani and B. Nail Chiwetalu (2006). "Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: the application of traditional and comprehensive Pearl Growing. A review." International journal of language and communication disorders **41**(5): 567-582.
- Schulman, J. L. (2011). "What's New for 2012 MeSH®." NLM Technical Bulletin **November-December** (383): e8.

- Shaikh, N., R. Badgett, M. Pi, N. Wilczynski, K. A. McKibbin, A. Ketchum and R. B. Haynes (2011). "Development and validation of filters for the retrieval of studies of clinical examination from Medline." Journal of Medical Internet Research **13**(4): e82-e82.
- Shekelle, P. G., P. J. Pronovost, R. M. Wachter, S. L. Taylor, S. M. Dy, R. Foy, S. Hempel, K. M. McDonald, J. Ovreteit, L. V. Rubenstein, A. S. Adams, P. B. Angood, D. W. Bates, L. Bickman, P. Carayon, L. Donaldson, N. Duan, D. O. Farley, T. Greenhalgh, J. Haughom, E. T. Lake, R. Lilford, K. N. Lohr, G. S. Meyer, M. R. Miller, D. V. Neuhauser, G. Ryan, S. Saint, K. G. Shojania, S. M. Shortell, D. P. Stevens and K. Walshe (2011). "Advancing the science of patient safety." Ann Intern Med **154**(10): 693-696.
- Shojania, K., B. Duncan, K. McDonald and R. Wachter (2002). "Safe but sound: patient safety meets evidence-based medicine." JAMA (Chicago, Ill.) **288**(4): 508-513.
- Shojania, K. G., B. W. Duncan, K. M. McDonald, R. M. Wachter and A. J. Markowitz (2001). "Making health care safer: a critical analysis of patient safety practices." Evid Rep Technol Assess (Summ)(43): i-x, 1-668.
- Sistrom, C. L. and C. W. Garvan (2004). "Proportions, Odds, and Risk1." Radiology **230**(1): 12-19.
- Smith, W. (1996). "The effects of base rate and cutoff point choice on commonly used measures of association and accuracy in recidivism research." Journal of Quantitative Criminology **12**(1): 83-111.
- Snavey, L. and N. Cooper (1997). "The information literacy debate." Journal of academic librarianship **23**(1): 9-14.
- Spiegel, A. D. and F. Kavalier (1997). "America's first medical malpractice crisis, 1835-1865." Journal of Community Health **22**(4): 283-308.
- Steinbrook, R. (2002). "Protecting Research Subjects - The Crisis at Johns Hopkins." New England Journal of Medicine **346**(9): 716-720.
- Stelfox, H. T., S. Palmisani, C. Scurlock, E. J. Orav and D. W. Bates (2006). "The "To Err is Human" report and the patient safety literature." Qual Saf Health Care **15**(3): 174-178.
- Straus, S. E. (2005). Evidence-based medicine : how to practice and teach EBM. Edinburgh ; New York, Elsevier/Churchill Livingstone.
- Straus, S. E. and F. A. McAlister (2000). "Evidence-based medicine: a commentary on common criticisms." Canadian Medical Association. Journal; CMAJ **163**(7): 837-841.
- Straus, S. E., W. S. Richardson, P. Glasziou and R. B. Haynes (2005). Evidence-based medicine : how to practice and teach EBM. Edinburgh ; New York, Elsevier/Churchill Livingstone.

- Stryker, J. E., R. J. Wray, R. C. Hornik and I. Yanovitzky (2006). "Validation of database search terms for content analysis: The case of cancer news coverage." Journalism & Mass Communication Quarterly **83**(2): 413-430.
- Su, L. T. (1993). "Is relevance an adequate criterion for retrieval system evaluation. An empirical inquiry into the user's evaluation." Proceedings of the 56th Annual Meeting of the American Society for Information Science: 93-103.
- Swanson, D. R. (1965). "The Evidence Underlying the Cranfield Results." The Library Quarterly **35**(1): 1-20.
- Swanson, E. B. (2009). Information Systems. Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 2635-2642.
- Tague-Sutcliffe, J. (2009). Information Retrieval Experimentation. Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 2526-2534.
- Tanon, A. A., F. Champagne, A. P. Contandriopoulos, M. P. Pomey, A. Vadeboncoeur and H. Nguyen (2010). "Patient safety and systematic reviews: finding papers indexed in MEDLINE, EMBASE and CINAHL." Qual Saf Health Care **19**(5): 452-461.
- Taylor, H. C. and J. T. Russell (1939). "The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables." Journal of Applied Psychology **23**(5): 565-578.
- Taylor, R. S. (1962). "The process of asking questions." American Documentation **13**(4): 391-396.
- Taylor, R. S. (1968). "Question-negotiation and information seeking in libraries." College & research libraries **29**(3): 178.
- The Cochrane Collaboration (2011). Review Manager (RevMan). Copenhagen: The Nordic Cochrane Centre,.
- Thibault, M., S. Bagel, V. Sautou-Miranda and J. Chopineau (2001). "Séries statistiques simples." Lyon pharmaceutique **52**: 214-232.
- Thomas, A., A. Saroyan and W. D. Dauphinee (2011). "Evidence-based practice: a review of theoretical assumptions and effectiveness of teaching and assessment interventions in health professions." Advances in health sciences education **16**(2): 253-276.
- Thomas, E. J., D. M. Studdert, H. R. Burstin, E. J. Orav, T. Zeena, E. J. Williams, K. M. Howard, P. C. Weiler and T. A. Brennan (2000). "Incidence and types of adverse events and negligent care in Utah and Colorado." Med Care **38**(3): 261-271.
- Thomas, E. J., D. M. Studdert, J. P. Newhouse, B. I. Zbar, K. M. Howard, E. J. Williams and T. A. Brennan (1999). "Costs of medical injuries in Utah and Colorado." Inquiry **36**(3): 255-264.

- Thomas, J., J. Brunton and S. Graziosi. (2010). "EPPI-Reviewer 4: software for research synthesis." Retrieved June 2012, from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2914>.
- Thomas, J., J. McNaught and S. Ananiadou (2011). Applications of text mining within systematic reviews. Research Synthesis Methods. **2**: 1-14.
- Thomas, M. and K. Mackway-Jones (2008). "Incidence and causes of critical incidents in emergency departments: a comparison and root cause analysis." Emergency Medicine Journal **25**(6): 346-350.
- Tuominen, K. and R. Savolainen (1997). A social constructionist approach to the study of information use as discursive action. Information seeking in context. Proceedings of an international conference on research in information needs, seeking and use in different contexts, 14-16 August 1996, Tampere, Finland. P. Vakkari, R. Savolainen and B. Dervin. London, Taylor Graham: pp. 81-96.
- U.S. National Library of Medicine - National Institutes of Health. (2011). "Medline: Number of Citations to English Language Articles; Number of Citations Containing Abstracts." Retrieved January 5, 2012, from http://www.nlm.nih.gov/bsd/medline_lang_distr.html.
- U.S. National Library of Medicine - National Institutes of Health. (2012). "Fact Sheet MEDLINE®." Retrieved June, 2012, from <http://www.nlm.nih.gov/pubs/factsheets/medline.html>.
- Ugolini, D., M. Neri, C. Casilli and S. Bonassi (2010). "Development of search filters for retrieval of literature on the molecular epidemiology of cancer." Mutation Research **701**(2): 107-110.
- Van den Heede, K., W. Sermeus, L. Diya, E. Lesaffre and A. Vleugels (2006). "Adverse outcomes in Belgian acute hospitals: retrospective analysis of the national hospital discharge dataset." International Journal for Quality in Health Care **18**(3): 211-219.
- Vincent, C., G. Neale and M. Woloshynowych (2001). "Adverse events in British hospitals: preliminary retrospective record review." Bmj **322**(7285): 517-519.
- Vincent, S., S. Greenley and O. Beaven (2003). "Clinical Evidence diagnosis: Developing a sensitive search strategy to retrieve diagnostic studies on deep vein thrombosis: a pragmatic approach." Health Info Libr J **20**(3): 150-159.
- Voorhees, E. M. (2009). Text REtrieval Conference (TREC). Encyclopedia of Library and Information Sciences, Third Edition, Taylor & Francis: 5182-5189.
- Walker, C. J., K. A. McKibbin, R. B. Haynes and M. F. Ramsden (1991). "Problems encountered by clinical end users of MEDLINE and GRATEFUL MED." Bulletin of the Medical Library Association **79**(1): 67-69.
- Warrens, M. (2008). "On Association Coefficients for 2x2 Tables and Properties That Do Not Depend on the Marginal Distributions." Psychometrika **73**(4): 777-789.

- Weber, M., J. A. Kors and B. Mons (2005). Online tools to support literature-based discovery in the life sciences, Oxford University Press / UK. **6**: 277-286.
- Weightman, A. L. and J. Williamson (2005). "The value and impact of information provided through library services for patient care: a systematic review." *Health Info. Libr. J.* **22**(1): 4-25.
- Weiss, N. A. and M. J. Hassett (1991). Introductory statistics. Reading, Mass., Addison-Wesley Pub. Co.
- Westwood, M., M. Rodgers, A. Sowden, J. F. Cozens, S. Golder and J. Glanville (2002). Patient safety: A Mapping of the Research Literature. York, University of York, NHS Centre for Reviews and Dissemination.
- White, V. J., J. M. Glanville, C. Lefebvre and T. A. Sheldon (2001). "A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy." *Journal of information science* **27**(6): 357-370.
- Whitford, A. B. (2005). Correlations. Encyclopedia of Social Measurement. K. Kempf-Leonard. New York, Elsevier: 523-529.
- Whiting, P., M. Westwood, R. Beynon, M. Burke, J. A. Sterne and J. Glanville (2011). "Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies." *J Clin Epidemiol* **64**(6): 602-607.
- Whiting, P., M. Westwood, M. Burke, J. Sterne and J. Glanville (2008). "Systematic reviews of test accuracy should search a range of databases to identify primary studies." *J Clin Epidemiol* **61**(4): 357-364.
- Wilczynski, N. L., A. X. Garg and B. Haynes (2007). "A method for defining a journal subset for a clinical discipline using the bibliographies of systematic reviews." *Stud Health Technol Inform* **129**(Pt 1): 721-724.
- Wilczynski, N. L., R. B. Haynes, J. N. Lavis, R. Ramkissoonsingh, A. E. Arnold-Oatley and A. E. (2004). "Optimal search strategies for detecting health services research studies in MEDLINE." *CMAJ* **171**(10): 1179-1185.
- Wong, S. S., N. L. Wilczynski and R. B. Haynes (2006). "Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE." *J Med Libr Assoc* **94**(4): 451-455.
- Woods, D. and K. Trewheellar (1998). "Medline and Embase complement each other in literature searches." *BMJ* **316**(7138): 1166.
- Wyer, P., T. Allen and C. J. Corral (2003). "How to find evidence when you need it, part 4: Matching clinical questions to appropriate databases." *Annals of emergency medicine* **42**(1): 136-149.
- Yang, Y., L. Akers, T. Klose and C. B. Yang (2008). "Text mining and visualization tools – Impressions of emerging capabilities." *World patent information* **30**(4): 280.

- Yasnoff, W. A., P. W. O'Carroll, D. Koo, R. W. Linkins and E. M. Kilbourne (2000). "Public health informatics: improving and transforming public health in the information age." J Public Health Manag Pract **6**(6): 67-75.
- Younger, P. and K. Boddy (2009). "When is a search not a search? A comparison of searching the AMED complementary health database via EBSCOhost, OVID and DIALOG." Health information and libraries journal **26**(2): 126-135.
- Zeller, R. A. (2005). Measurement Error, Issues and Solutions. Encyclopedia of Social Measurement. K. Kempf-Leonard. New York, Elsevier: 665-676.
- Zipperer, L. (2004). "Clinicians, librarians and patient safety: opportunities for partnership." Qual Saf Health Care **13**(3): 218-222.
- Zlowodzki, M., B. A. Zelle, M. Keel, P. A. Cole and P. J. Kregor (2006). "Evidence-based resources and search strategies for orthopaedic surgeons." Injury **37**(4): 307-311.

Annexe 1 : Sélection de journaux pour la recherche manuelle

Tableau _annexe 1 : Application des trois premiers critères de sélection :

Titre du journal	Répertorié dans une base de donnée (1=Oui, 0=Non)			Nbre de N°/an	Disponibilité en format électronique	
	Medline	Embase	Cinhal			À partir de
Academic Emergency Medicine	1	1	1	12	Oui	1994
Academic Medicine	1	1	1	12	Oui	2000
Accident analysis & prevention	1	1	0	6	Oui	1985
American journal of critical care	1	0	1	6	Oui	1992
American journal of emergency medicine	1	1	1	7	Oui	1985
American journal of epidemiology	1	1	1	24	Oui	1985
American journal of managed care	1	1	1	12	Oui	1995
American journal of medicine	1	1	1	12	Oui	1985
American journal of medical quality	1	0	1	6	Oui	1998
American journal of nursing	1	1	1	12	Oui	1985
American journal of preventive medicine	1	1	0	8	Oui	1998
American journal of public health	1	1	1	12	Oui	1985
American journal of surgery	1	1	0	12	Oui	1985
Annals of emergency medicine	1	1	1	12	Oui	1985
BMC-Health services research	1	1	0	12	Oui	2001
British journal of nursing	1	0	1	24	Oui	1992
British medical journal	1	1	1	52	Oui	1994
Canadian medical association journal	1	1	1	24	Oui	1985
Canadian journal of surgery	1	1	1	6	Oui	1985
Clinical biochemistry	1	1	0	12	Oui	1985
Critical care clinics	1	1	1	4	Oui	1998
Critical care medicine	1	1	1	12	Oui	1995
Effective clinical practice	1	0	0	12	Oui	1998
Family practice	1	1	0	6	Oui	1985
Frontiers of health services management	1	1	0	4	Oui	1997
Health psychology	1	1	1	6	Oui	1993
Health services research	1	1	1	8	Oui	2002
Hong-Kong medical journal	1	1	0	4	Oui	1995
International journal of clinical practice	1	1	1	8	Oui	2004
International journal of health services	1	1	1	4	Oui	1985
International journal of nursing practice	1	0	1	6	Oui	1995
International journal of nursing studies	1	0	1	6	Oui	1985
International journal for quality in health care	1	1	0	6	Oui	1989
Journal of clinical epidemiology	1	1	0	12	Oui	1988
Journal of clinical nursing	1	0	1	6	Oui	1997
Journal of critical care	1	1	1	4	Oui	1986
Journal of evaluation in clinical practice	1	1	0	4	Oui	1997
Journal of health economics	1	1	0	6	Oui	1985
Journal of intensive care medicine	1	1	1	6	Oui	1987
Journal of nursing administration	1	1	1	11	Oui	1996

Titre du journal	Répertorié dans une base de donnée (1=Oui, 0=Non)			Nbre de N°/an	Disponibilité en format électronique	
	Medline	Embase	Cinhal		À partir de	
Journal of quality in clinical practice	1	1	0	4	Oui	1998
Journal of safety research	1	1	0	5	Oui	1985
Journal of the American medical association	1	1	1	48	Oui	1985
Joint commission journal on quality and safety	1	0	1	12	Oui	2000
Lancet journal	1	1	1	48	Oui	1985
Medical journal of Australia MJA	1	1	1	12	Oui	1996
Milbank Quarterly	1	1	0	4	Oui	1997
New England journal of medicine	1	1	1	48	Oui	1985
Nursing	1	0	1	12	Oui	1985
Nursing management	1	0	1	12	Oui	2001
Nursing research	1	1	1	6	Oui	1985
Patient education and counselling	1	1	1	6	Oui	1985
Psychological medicine	1	1	0	12	Oui	1997
Quality management in health care journal	1	0	1	4	Oui	1999
Quality & safety in health care journal	1	1	1	24	Oui	1992

La syntaxe utilisée dans Medline pour la sélection définitive des journaux pour la recherche manuelle.

MEDLINE OVID 1950-2007

#1	Medical Errors/
#2	Safety/
#3	Explode Patients/
#4	#2 and #3
#5	#1 or #4

Annexe 2 : Les formulaires développés pour guider le choix des articles du gold standard.

Formulaire pour la première étape (Titre et Résumé).

Formulaire développé pour la première étape de sélection des articles du gold standard.	
Objectif de la première étape: À partir du titre et du résumé, faire un premier tri pour les articles susceptibles de nous intéresser dans le cadre de la revue.	
Contribution: Cette étape contribuera aux résultats attendus suivants: Repérer les mots libres couramment utilisés dans les textes pour parler du concept « Patient Safety»; Identifier un corpus d'articles à partir desquels évaluer la qualité des stratégies de recherche.	
Est ce qu'on peut dire que l'article traite de la Sécurité des Patients dans les organisations de Santé?	
	<input type="radio"/> Oui
	<input type="radio"/> Non
	<input type="radio"/> Ne peut me prononcer
Est ce qu'on retrouve dans le titre et/ou le résumé une expression faisant penser à la SÉCURITÉ des PATIENTS dans les ORGANISATIONS DE SANTÉ (Cochez toutes celles qui s'appliquent)?	
	<input type="checkbox"/> Patient Safety
	<input type="checkbox"/> Medical Error(s)
	<input type="checkbox"/> Medication Error(s)
	<input type="checkbox"/> Diagnostic Error(s)
	<input type="checkbox"/> Adverse event
	<input type="checkbox"/> Adverse effect
	<input type="checkbox"/> Adverse outcome
	<input type="checkbox"/> Undesirable event
	<input type="checkbox"/> Undesirable effect

	<input type="checkbox"/>	Undesirable outcome	
	<input type="checkbox"/>	Medical injury	
	<input type="checkbox"/>	Other expressions (Please specify)	
	<input type="checkbox"/>	No expressions	

Formulaire pour la deuxième étape (Textes complets).

Formulaire développé pour la deuxième étape de sélection des articles du gold standards	
Objectifs de la deuxième étape: À partir des textes complets, faire une sélection définitive des articles susceptibles de nous intéresser dans le cadre de la revue.	
Contribution: Cette étape contribuera aux résultats attendus suivants: <ul style="list-style-type: none"> • Repérer les mots libres couramment utilisés dans les textes pour parler du concept « Patient Safety»; • Repérer les mots libres qui nous permettront d'identifier les articles appropriés pour une analyse de concept; • Identifier un corpus d'articles à partir desquels évaluer la qualité des stratégies de recherche 	
1. Après avoir lu le texte complet, est-ce on peut dire que l'article traite de la <u>Sécurité des Patients</u> dans les <u>Organisations de Santé</u>?	
	<input type="radio"/> Oui <input type="radio"/> Non
2. Quelles sont les mots ou expressions que l'on retrouve dans les sections TITRE, RÉSUMÉ, et KEYWORDS qui font penser à la SÉCURITÉ des PATIENTS dans les ORGANISATIONS DE SANTÉ (Cochez toutes celles qui s'appliquent)?	
	<input type="checkbox"/> Patient Safety
	<input type="checkbox"/> Medical Error(s)
	<input type="checkbox"/> Medication Error(s)
	<input type="checkbox"/> Diagnostic Error(s)
	<input type="checkbox"/> Adverse event
	<input type="checkbox"/> Adverse effect
	<input type="checkbox"/> Adverse outcome
	<input type="checkbox"/> Undesirable event
	<input type="checkbox"/> Undesirable effect
	<input type="checkbox"/> Undesirable outcome
	<input type="checkbox"/> Medical injury

	<input type="checkbox"/>	Other expressions (Please specify)	
	<input type="checkbox"/>	No expressions	

Annexe 3 : Exemple de sélection de termes de recherche.

Les termes commençant par la Lettre S dans les Titres et les Résumés des articles composant le gold standard Composition dans Medline.

LA LETTRE S

safe	SEIPS	show	sophisticated	STRATEGIES	suffer
safer	select	showed	sought	strategies,	sufficient
safer.	selected	shown	source	strategies.	sufficiently
Safety	selecting	shows	sources,	STRATEGY	suggest
safety"	selection	sick	SP	strategy.	suggested
safety)	selective	side	spawning	streamline	suggestions
safety,	seminars	signal	speak	strength	suggests
Safety.	semi-structured	signals,	speciality,	Strengthening	suicides
safety:	send	signals.	specialties	strengths	suitable
safety;	Senior	Significant	specialty,	strides	supervising
safety?	sense	significant.	specific	striking	supervisors
safety-driven	sensitivity	significantly	specified	striving	supervisory
sample	separate	Similar	spends	Strong	supplemented
sample.	series	similarities	spinal	structural	support
sample:	serious	similarly	spread	structure	support,
satisfaction	serious,	simple	SPs	structured	support.
satisfactory	seriousness	simulation	St	structures,	Surgeons
satisfied	seriousness,	simulation,	Staff	student	surgeons'
Scale	seriousness.	simultaneously.	staff,	students	surgeons,
scale)	service	Since	staffing	Studies	surgery
scale.	services	Singapore	staffing,	studies,	surgery)
scales,	serving	single	stage	studies.	surgical
scenario	Session	situation	stage,	Study	surrounding
scenario-	set	situation.	stakeholders	study,	surveillance
assignment	sets,	Situational	standard	study.	Survey
scenarios	SETTING	situations	standard"	styles	survey.
scientific	SETTING,	situations,	standardisation	sub-categories	surveys
SCM	setting.	situations.	Standardised	subgroup	surveys.
scope	SETTING:	six	standardization.	subject	survival
score	settings	Sixty	standards	subjective	suspected
scores	settings,	skewness	standards.	SUBJECTS	sustain
scores.	settings.	skill	starting	subjects,	syndrome",

LA LETTRE S

screened	settle	skilled	State	submitted	SynergyHealth,
screening	settlement	skills	stated	subpractices,	synthesizes
scrutinized,	Seven	slightly	States	subscale	System
SCSc	Several	Slips,	States.	subsequent	system,
SCSc.	severe	slow	statewide	subset	system.
SCSu	severity	small	statistic.	substandard	system;
SCSu,	severity,	smaller	statistically	Substantial	Systematic
se	sex	Smart	statistics	substantially	systematically
search	sex,	so	statistics,	subtype	Systems
searched	share	so.	stay	success	system's
second	shared	social	stem	successful	systems,
second.	sharing	societal	steps:	successfully	systems.
secondly,	sheets,	software	still	such	
sectional	shift	solicitation	stimulated		
sectors.	Shojania	solicited	strategic		
security	shortcomings	solutions			
seeking	short-term	solving,			
seemed	should	Some			
seen		sometimes			

Préparation d'un terme sélectionné commençant par S pour interroger la base Medline.

Pour le terme Safety, on retrouve les mots suivants dans le titre et le résumé:

Safe; safer; safer.; Safety; safety"; safety); safety,; Safety.; safety.; safety;; safety?; safety-driven

Pour les retrouver dans Medline les requêtes suivantes seront testées:

Safe\$.ti,ab.

Safe\$3.ti,ab.

Annexe 4 : The original strategies used by Westwood et al. and their translation into the OVID and EBSCOhost interfaces.

Presentation¹⁰

The Westwood Study published in 2002 was designed to draw together a diverse collection of research and provide a map of the patient safety research literature. In particular the following questions were addressed:

1. What have been the goals of patient safety research?
2. What methods have been used in patient safety research?
3. What types of studies have shown what kinds of results?

Studies included were those in which patient safety falls somewhere on the continuum from what have been called errors and deviations, through dangerous situations and near misses to accidents. Research studies that have investigated the roots of individual or system failure, detection of error and direct interventions designed to reduce accidents have been included. Hospital care has been the main focus (as this is where the majority of research has been conducted) although other health care settings have been considered.

Efforts to improve safety require an understanding of the organisational culture and working conditions that influence safe conduct, therefore research studies addressing this issue have also been included. All types of research have been considered, both quantitative and qualitative to reflect the approaches used in patient safety research. Publications that

¹⁰ Copied from the Westwood Study report: Westwood M, Rodgers M, Sowden A. Patient safety: a mapping of the research literature. York, UK: NHS Centre for Reviews and Dissemination.

are potentially of interest but which fall outside of being empirical research have not been included, due to constraints of time. In particular there are a number of key texts, for example: *Clinical Risk management: Enhancing Patient Safety*, edited by Vincent⁶ which presents useful material, ranging from the principles of risk management through to the implementation of risk management strategies, which haven't been included as they present little if any direct research. The searches have dated back to 1990, as this is when a body of research documenting the problem of medical errors began to emerge.

The strategies

INITIAL STRATEGY MEDLINE (Silverplatter/ARC searched – 03/07/01) 1966 – 2000/12	OUR TRANSLATION MEDLINE OVID Gateway1950-2008
1. "Accident-Prevention"	1 Accident Prevention/
2. "Safety"/ all subheadings	2 Safety/
3. "Accident-Proneness"	3 Accident Proneness/
4. "Medical-Errors"/ all subheadings	4 Medical Errors/
5. "Medication-Errors"/ all subheadings	5 Medication Errors/
6. patient safety in ti ab	6 patient safety.ti,ab.
7. medical error* in ti ab	7 medical error\$.ti,ab.
8. medication error* in ti ab	8 medication error\$.ti,ab.
9. human error* in ti ab	9 human error\$.ti,ab.
10. detect* near ((error* or incident* or accident*) in ti ab)	10 (detect\$ adj (error\$ or incident\$ or accident\$)).ti,ab.
11. reporting near ((error* or incident* or accident* or system*) in ti ab)	11 (reporting adj (error\$ or incident\$ or accident\$)).ti,ab.
12. information system* in ti ab	12 information system\$.ti,ab.
13. prescribing system* in ti ab	13 prescribing system\$.ti,ab.
14. "Organizational-Culture"	14 Organizational Culture/
15. "Fatigue"/ all subheadings	15 Fatigue/
16. explode "Stress-Psychological"/ all subheadings	16 exp Stress, Psychological/
17. explode "Information-Systems"/ all subheadings	17 exp Information Systems/
18. "Risk-Management"/ all subheadings	18 Risk Management/
19. "Safety-Management"/ all subheadings	19 Safety Management/
20. Risk Management in ti ab	20 risk management.ti,ab.
21. Safety Management in ti ab	21 safety management.ti,ab.
22. (fatigue or workload* or tired*) near ((error* or harm or incident* or accident* or patient safety or job performance or work performance) in ti ab)	22 ((fatigue or workload\$ or tired\$) adj (error\$ or harm or incident\$ or accident\$ or patient safety or job performance or work performance)).ti,ab.
23. incident* report* in ti ab	23 incident\$ report\$.ti,ab.
24. #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 #8 or #9	24 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 9
25. #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9	25 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9
26. #25 and #24	26 24 and 25

INITIAL STRATEGY EMBASE (Silverplatter/ARC – searched - 06/07/01) 1980 – 2001/06	OUR TRANSLATION EMBASE OVID Gateway1996-2008
1. "Medical-Errors"/ all subheadings	1 Medical Error/
2. "Medication-Errors"/ all subheadings	2 Medication Error/
3. patient safety in ti ab	3 patient safety.ti,ab.
4. medical error* in ti ab	4 medical error\$.ti,ab.
5. medication error* in ti ab	5 medication error\$.ti,ab.
6. human error* in ti ab	6 human error\$.ti,ab.
7. detect* near ((error* or incident* or accident*) in ti ab)	7 (detect\$ adj (error\$ or incident\$ or accident\$)).ti,ab.
8. reporting near ((error* or incident* or accident* or system*) in ti ab)	8 (reporting adj (error\$ or incident\$ or accident\$ or system\$)).ti,ab.
9. information system* in ti ab	9 information system\$.ti,ab.
10. prescribing system* in ti ab	10 prescribing system\$.ti,ab.
11. "Risk-Management"/ all subheadings	11 risk management/
12. Risk Management in ti ab	12 Risk Management.ti,ab.
13. Safety Management in ti ab	13 Safety Management.ti,ab.
14. (fatigue or workload* or tired*) near ((error* or harm or incident* or accident* or patient safety or job performance or work performance) in ti ab)	14 ((fatigue or workload\$ or tired\$) adj (error\$ or harm or incident\$ or accident\$ or patient safety or job performance or work performance)).ti,ab.
15. incident* report* in ti ab	15 incident\$ report\$.ti,ab.
16. ("Accident-Prevention") or ("Safety"/ all subheadings) or ("Accident-Proneness") or #1 or #2 or #3 or #4 or #5 or #6	16 Accident Prevention/ or Safety/ or Accident Proneness/ or 1 or 2 or 3 or 4 or 5 or 6
17. #7 or #8 or #9 or #10 or ("Fatigue"/ all subheadings) or #11 or #12 or #13 or #14 or #15 or (explode "stress"/ all subheadings)	17 7 or 8 or 9 or 10 or Fatigue/ or 11 or 12 or 13 or 14
18. #1 or #2 or #3 or #4 or #5 or #6	18 1 or 2 or 3 or 4 or 5 or 6
19. #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15	19 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15
20. #18 and #19	20 18 and 19

INITIAL STRATEGY CINAHL (Silverplatter/ARC – searched - 06/07/01) 1982 – 2001/04	OUR TRANSLATION CINAHL EBSCOhost 1982-2008)
1. "Patient-Safety"/ all topical subheadings / all age subheadings	S1 (MH "Patient Safety")
2. explode "Treatment-Errors"/ all topical subheadings / all age subheadings	S2 (MH "Treatment Errors+")
3. explode "Diagnostic-Errors"/ all topical subheadings / all age subheadings	S3 (MH "Diagnostic Errors+")
4. patient safety in ti ab	S4 TI (patient safety) OR AB (patient
5. medical error* in ti ab	safety)
6. detect* near ((error* or incident* or accident*) in ti ab)	S5 TI (medical error*) OR AB
7. reporting near ((error* or incident* or accident* or system*) in ti ab)	(medical error*)
8. information system* in ti ab	S6 TI (detect* AND (error* OR
9. prescribing system* in ti ab	incident* OR accident*)) OR AB (detect* AND
10. "Organizational-Culture"	(error* OR incident* OR accident*))
11. "Fatigue"/ all subheadings/ all age subheadings	S7 TI (reporting AND (error* OR
12. explode "Stress-Psychological"/ all subheadings/ all age subheadings	incident* OR accident* OR system*)) OR AB
13. explode "Information-Systems"/ all subheadings/ all age subheadings	(reporting AND (error* OR incident* OR accident*
14. "Risk-Management"/ all subheadings/ all age subheadings	OR system*))
15. Safety Management in ti ab	S8 TI (information system*) OR AB
16. (fatigue or workload* or tired*) near ((error* or harm or incident* or accident* or patient safety or job performance or work performance) in ti ab)	(information system*)
17. incident* report* in ti ab	S9 TI (prescribing system*) OR AB
18. #1 or #2 or #3 or #4 or #5	(prescribing system*)
19. #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17	S10 (MH "Organizational Culture")
20. #18 and #19	S11 (MH "Fatigue")
	S12 (MH "Stress, Psychological+")
	S13 (MH "Information Systems+")
	S14 (MH "Risk Management")
	S15 TI (Safety management) OR AB
	(Safety management)
	S16 TI ((fatigue or workload* or
	tired*) AND (error* or harm or incident* or
	accident* or (patient AND safety) or (job AND
	performance) or (work AND performance))) or AB (
	(fatigue or workload* or tired*) AND (error* or
	harm or incident* or accident* or (patient AND
	safety) or (job AND performance) or (work AND
	performance)))
	S17 TI (incident* report*) OR AB
	(incident* report*)
	S18 S1 OR S2 OR S3 OR S4 OR S5
	S19 S6 OR S7 OR S8 OR S9 OR S10
	OR S11 OR S12 OR S13 OR S14 OR S15 OR S16
	OR S17
	S20 S18 OR S19

Annexe 5 : Search strategies to retrieve patient safety papers in MEDLINE, EMBASE and CINAHL.

MEDLINE		EMBASE		CINAHL	
<p>S₁-MED (exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit/)</p>	<p>Sensitivity 100.00% (100%)⁴</p>	<p>S₁-EMB (Exp Safety/ OR Safe\$.ti,ab. OR Err\$.ti,ab. OR Adverse.ti,ab) AND [(Exp Health Care Quality/ OR Risk assessment/) OR (exp *Safety/ AND *Medical Practice)]</p>	<p>Sensitivity 100.00% (95.00%)</p>	<p>S₁-CIN [(Tx Health* OR Tx Care* OR Tx Prevent*) AND (Tx Safe* OR Tx Err*) AND (MH "Child" OR Tx Patient* or Tx Human)] OR MH "Health Care Errors+".</p>	<p>Sensitivity 100.00% (100%)</p>
<p>S₂-MED (safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.) AND (exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit/)</p>	<p>Sensitivity 98.80% (95.31%)</p>	<p>S₂-EMB (Exp Safety/ OR Safe\$.ti,ab. OR Err\$.ti,ab. OR Adverse.ti,ab) AND [exp Health Care Quality/ OR ((Exp Safety/ OR Err\$.ti,ab.) AND (Exp Health Care Organization/ OR Exp Health Care/ OR Medic\$.ti,ab. OR Exp Health Care System/ OR Hospital\$.ti,ab.))]</p>	<p>Sensitivity 98.73% (95.00%)</p>	<p>S₂-CIN (Tx Health* or Tx Care* or Tx Prevent*) AND (Tx Safe* or Tx Err*)</p>	<p>Sensitivity 98.21% (96.15%)</p>
<p>S₃-MED (safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.)</p>	<p>Sensitivity 98.80% (95.31%)</p>	<p>S₃-EMB (exp Medical Error/ OR Patient Safety/ OR Exp Iatrogenic disease/) OR [(Exp Safety/ OR Safe\$.ti,ab. OR Err\$.ti,ab. OR Adverse.ti,ab.) AND [exp Health Care Quality/ OR (exp *Safety/ AND (Exp Health Care Organization/ OR Exp Health Care/ OR Medic\$.ti,ab. OR Exp Health Care System/ OR Hospital\$.ti,ab.))]]</p>	<p>Sensitivity 94.94% (93.33%)</p>	<p>S₃-CIN [(Tx Health* OR Tx Care* OR Tx Prevent*) AND (Tx Safe* OR Tx Err*)] AND (MH "Child" OR Tx Patient* OR Tx Human).</p>	<p>Sensitivity 96.43% (86.54%)</p>
<p>P₁-MED *Medical Errors/</p>	<p>Precision⁵ 51.35% (36.64%)</p>	<p>P₁-EMB (Exp Safety/ OR Err\$.ti,ab.) AND (patient\$.ti,ab.) AND (Exp Health Care Organization/ OR Exp Health Care/ OR Medic\$.ti,ab. OR Exp Health Care System/ OR Hospital\$.ti,ab.)</p>	<p>Precision 45.61% (45.71%)</p>	<p>P₁-CIN Tx Harm* AND Tx Event*</p>	<p>Precision 61.54% (83.33%)</p>
P₂-MED	Precision	P₂-EMB	Precision	P₂-CIN	Precision

MEDLINE		EMBASE		CINAHL	
(exp *Medical Errors/ OR *Safety Management/) AND (patient\$.ti,ab.)	51.35% (32%)	(Exp Safety/ OR Err\$.ti,ab.) AND [(Patient Care/ OR (patient\$.ti,ab.) AND (Exp Health Care Organization/ OR Exp Health Care/ OR Medic\$.ti,ab. OR Exp Health Care System/ OR Hospital\$.ti,ab.))]	44.92% (45.71%)	Tx Err* AND Tx Event*	55.00% (59.38%)
P₃-MED (exp *Risk Management/) AND (patient\$.ti,ab.)	Precision 51.02% (34.18%)	P₃-EMB (Exp Safety/ OR Err\$.ti,ab.) AND [(Patient Care/ OR (patient\$.ti,ab. OR Human/ OR Human.ti,ab.) AND (Exp Health Care Organization/ OR Health\$.ti,ab.))]	Precision 43.20% (48.53%)	P₃-CIN [(Tx Harm* AND Tx Event*) OR MH "Risk Management+"] OR (MH "Child" OR Tx Patient* OR Tx Human)	Precision 50.00% (52.78%)
SP₁-MED	Sensitivity * Precision⁶ 41.05% (30.60%)	SP₁-EMB [(Exp Safety/ OR Err\$.ti,ab.) AND (Health Care Quality)] OR [(Exp Safety/ OR Safe\$.ti,ab. OR Err\$.ti,ab. OR Adverse.ti,ab.) AND (Exp Health Care Organization/ OR Health\$.ti,ab.)]	Sensitivity * Precision 33.49% (33.07%)	SP₁-CIN MH "Health Care Errors+"	Sensitivity * Precision 32.47% (35.11%)
SP₂-MED (exp *Medical Errors/ OR *Safety Management/) AND (safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.)	Sensitivity * Precision 44.31% (32.09%)	SP₂-EMB (Exp Safety/ OR Safe\$.ti,ab. OR Err\$.ti,ab. OR Adverse.ti,ab.) AND [(Health Care Quality/ OR ((Exp Safety/ OR Err\$.ti,ab.) AND (Exp Health Care Organization/ OR Health\$.ti,ab.))]	Sensitivity * Precision 32.82% (33.09%)	SP₂-CIN [([Tx Err* AND (Tx Health* OR Tx Care* OR Tx Prevent*)] OR MH "Risk Management+") AND (MH "Child" OR Tx Patient* OR Tx Human)] OR MH "Health Care Errors+".	Sensitivity * Precision 31.17% (33.13%)
SP₃-MED (safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.) AND (exp *Risk Management/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit)	Sensitivity * Precision 95.18% (95.31%)	SP₃-EMB (Exp Safety/ OR Safe\$.ti,ab. OR Err\$.ti,ab. OR Adverse.ti,ab.) AND [Health Care Quality OR (Exp Health Care Organization/ OR Health\$.ti,ab.)]	Sensitivity * Precision 31.89% (31.00%)	SP₃-CIN [([Tx Err* AND (Tx Health* OR Tx Care* OR Tx Prevent*)] OR MH "Risk Management+") AND (MH "Child" OR Tx Patient* OR Tx Human)]	Sensitivity * Precision 30.47% 31.55%

¹CGS (Composition Gold standard): a set of relevant records used to build and test the search strategies ; ²VGS (Validation Gold standard): a set of relevant records, different from CGS, used to test the search strategies and assess internal validity; ³(NUMBER): The performance of the search strategy in the Validation Gold Standard;

⁴SENSITIVITY: (Number of relevant articles retrieved by the search) / (TOTAL number of relevant articles in the universe of articles (gold standard)) * 100.;

⁵PRECISION: (Number of relevant articles retrieved by the search) / (TOTAL number of articles extracted by the search) * 100 ; ⁶SENSITIVITY*PRECISION: the product of sensitivity and precision, which allows us to reach an equilibrium between sensitivity and precision

Annexe 6 : Number and percentage of papers indexed with the new Mesh “Patient Safety” retrieved by the bibliographic search filters developed for Medline.

MeSH HEADING: PATIENT SAFETY

SCOPE: Efforts to reduce risk, to address and reduce incidents and accidents that may negatively impact healthcare consumers.

YEAR of ENTRY: 2012

Used For:

patient safety

patient safeties

safety patient

safeties patient

	April 17, 2012		October 20, 2012	
	Number of documents	% retrieved	Number of documents	% retrieved
Mesh Patient Safety	552	100,00%	1602	100,00%
S1 MED	403	73,01%	1252	78,15%
S2 MED	403	73,01%	1252	78,15%
S3 MED	552	100,00%	1602	100,00%
P1 MED	49	8,88%	126	7,87%
P2 MED	73	13,22%	198	12,36%
P3 MED	45	8,15%	101	6,30%
SP1 MED	106	19,20%	295	18,41%
SP2 MED	94	17,03%	272	16,98%
SP3 MED	143	25,91%	430	26,84%

Annexe 7 : Selection guides used to select the gold standard papers..

Selection guides used to select the gold standard papers.

Form developed for the first step of the gold standard papers selection process	
First step objectives: From the titles and abstracts, please make a first selection of papers that may be interesting for the review.	
Contribution: This first step will contribute to the following outcomes: <ul style="list-style-type: none"> • Identify text terms commonly used to discuss the "Patient Safety" concept. • Identify a corpus of articles that will be used to assess the quality the bibliographic search filters developed. 	
1. Can the paper examined be labeled as dealing with Patient safety in health Care organizations?	
	<input type="radio"/> Yes
	<input type="radio"/> No
	<input type="radio"/> Cannot tell
2. Is there any term in the TITLE or the ABSTRACT of the paper that suggests that the paper is dealing with Patient safety in Health care organizations? (Please select all that apply)?	
	<input type="checkbox"/> Patient Safety
	<input type="checkbox"/> Medical Error(s)
	<input type="checkbox"/> Medication Error(s)
	<input type="checkbox"/> Diagnostic Error(s)
	<input type="checkbox"/> Adverse event
	<input type="checkbox"/> Adverse effect
	<input type="checkbox"/> Adverse outcome
	<input type="checkbox"/> Undesirable event
	<input type="checkbox"/> Undesirable effect
	<input type="checkbox"/> Undesirable outcome

	<input type="checkbox"/>	Medical injury
	<input type="checkbox"/>	Other expressions (Please specify) <input type="text"/>
	<input type="checkbox"/>	No expressions
<input type="text"/>		

Form developed for the second step of the gold standard papers selection process

Second step objectives:

From the full text of the paper, please make a final selection of papers that may be interesting for the review.

Contribution:

This second step will contribute to the following outcomes:

- Identify text terms commonly used to discuss the "Patient Safety" topic.
- Identify text terms used to discuss conceptual aspects of "Patient Safety"
- Identify a corpus of articles that will be used to assess the quality the bibliographic search filters developed.

3. *After reading the full text of the paper, can the paper examined be labeled as dealing with Patient safety in health Care organizations?*

	<input type="radio"/>	Yes
	<input type="radio"/>	No

4. *What are the terms in the full text that suggest that the paper is dealing with Patient safety in Health care organizations? (Please select all that apply)?*

	<input type="checkbox"/>	Patient Safety	
	<input type="checkbox"/>	Medical Error(s)	
	<input type="checkbox"/>	Medication Error(s)	
	<input type="checkbox"/>	Diagnostic Error(s)	
	<input type="checkbox"/>	Adverse event	
	<input type="checkbox"/>	Adverse effect	
	<input type="checkbox"/>	Adverse outcome	
	<input type="checkbox"/>	Undesirable event	
	<input type="checkbox"/>	Undesirable effect	
	<input type="checkbox"/>	Undesirable outcome	
	<input type="checkbox"/>	Medical injury	
	<input type="checkbox"/>	Other expressions (Please specify)	<input style="width: 100%;" type="text"/>
	<input type="checkbox"/>	No expressions	

Annexe 8 : Caractéristiques des articles du gold standard NPSF en comparaison de ceux du gold standard qui a servi à la construction des filtres de recherche.

Document type	NPSF GOLD STANDARD			COMPOSITION GOLD STANDARD		
	Answers	% Answers	% Cases N= 3247	Answers	% Answers	% Cases N = 85
Journal Article	2950	75.9%	90.9%	71	65.1	83.5
Review	463	11.9%	14.3%	15	13.8	17.6
Editorial	145	3.7%	4.5%	10	9.2	11.8
News	81	2.1%	2.5%	1	0.9	1.2
Letter	39	1.0%	1.2%	3	2.8	3.5
Historical Article	26	0.7%	0.8%			
Interview	16	0.4%	0.5%			
Lectures	4	0.1%	0.1%			
Clinical Conference	7	0.2%	0.2%			
Comment	122	3.1%	3.8%	10	8.3	11.8
Congresses	12	0.3%	0.4%			
Biography	8	0.2%	0.2%			
Consensus Development Conference	5	0.1%	0.2%			
Guideline	7	0.2%	0.2%			
TOTAL	3885	100,0%	119,6%	109	100.0%	128.2%
Countries	Effectif	%	% Cumulatif	Effectif	%	% Cumulatif
USA	2557	73.6	73.6	52	61.2	61.2
UK, Canada and Australia	595	17.1	90.7	25	29.4	90.6
Other countries	322	9.1	100	8	9.4	100.0
TOTAL	3474	100.0%		85	100.0	
Institutions	Effectif	%	% Cumulatif	Effectif	%	% Cumulatif
Universities	1506	43.9	43.9	37	43.5	43.5
Health care organizations	955	27.8	71.7	24	28.2	71.7
Universities and health care organizations	217	6.3	78	8	9.4	81.1
Medical and Health professions Societies	68	2.0	80			
Private firms	142	4.1	84.1	3	3.5	84.6
Governmental agencies	155	4.5	88.6	5	5.9	90.5

	NPSF GOLD STANDARD			COMPOSITION GOLD STANDARD		
Journals	233	6.8	95.4	5	5.9	96.4
Think Thank and professionals associations	114	3.3	98.8	2	2.4	98.8
Patient Associations	14	0.4	99.2			
Mixte	28	0.8	100.0	1	1.2	100.0
TOTAL	3432	100.0		85	100.0	
Source	Effectif	%	% Cumulatif	Effectif	%	% Cumulatif
From academic journals	3148	89.7	89.7	85	100.0	100.0
From non academic sources	362	10.3	100.0			
TOTAL	3510	100.0				

Annexe 9 : Sélection des articles repérés par les filtres de recherche à l'aide d'un tirage aléatoire pour l'estimation de la précision dans le gold standard NPSF.

MEDLINE

S1-MED (<i>exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit/</i>)		
ANNÉE - (NUMÉRO DE LA REQUÊTE)	Total repéré dans OVID MEDLINE	Syntaxe
1999 (8)	136180	From 8 keep 99154, 13868, 92462, 45519, 128774, 119524, 104112, 104959, 89446, 20089
2000 (9)	143134	From 9 keep 18920, 127440, 59218, 106703, 142969, 131327, 140582, 32071, 54399, 2185
2001 (10)	152250	From 10 keep 99409, 96144, 38697, 142880, 57247, 43158, 92193, 132558, 138975, 20388
2002 (11)	163802	From 11 keep 35168, 155577, 78615, 54844, 146750, 2155, 17124, 133505, 38669, 154238
2003 (12)	180869	From 12 keep 121068, 171622, 118143, 115132, 140960, 81566, 10484, 90579, 141162, 106287
2004 (13)	196008	From 13 keep 112870, 33933, 111709, 186458, 187517, 70510, 101286, 155367, 64605, 152619
2005 (14)	210599	From 14 keep 14220, 44730, 159959, 88542, 55022, 43860, 152711, 108949, 50350, 126095
2006 (15)	220416	From 15 keep 163800, 20063, 13772, 8729, 182411, 83942, 203088, 206404, 38540, 69665
2007 (16)	233899	From 16 keep 16938, 34129, 209362, 59824, 190756, 189888, 199335, 30055, 169334, 179196
2008 (17)	245843	From 17 keep 171666, 222551, 31891, 25753, 96013, 118347, 191237, 104514, 56754, 21771
TOTAL	1883000	

S2-MED: (<i>safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.) AND (exp Risk Management/ OR exp Quality of Health Care/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit/)</i>)		
ANNÉE - (NUMÉRO DE LA REQUÊTE)		Syntaxe
1999 (7)	21239	From 7 keep 2228, 7078, 17901, 18681, 1086, 18326, 4742, 7581, 11754, 5304
2000 (8)	23908	From 8 keep 19565, 15533, 20728, 6074, 4804, 2900, 20196, 1881, 22993, 2927
2001 (9)	25888	From 9 keep 18087, 11344, 23880, 21664, 17381, 8699, 25433, 14745, 22173, 23550
2002 (10)	28261	From 10 keep 6111, 11134, 20990, 280, 1086, 16733, 7590, 26171, 7710, 13670
2003 (11)	31104	From 11 keep 30905, 21912, 14359, 15858, 3120, 23173, 8410, 9694, 3950, 11440
2004 (12)	34035	From 12 keep 24748, 14120, 7857, 25245, 14744, 49, 5618, 16307, 18161, 12963
2005 (13)	37252	From 13 keep 24497, 34023, 8969, 28700, 7409, 36892, 18586, 12707, 36019, 32367
2006 (14)	39787	From 14 keep 18514, 32182, 16916, 12481, 38341, 2679, 21569, 21681, 26038, 39326
2007 (15)	42817	From 15 keep 19317, 30980, 32717, 19003, 5205, 21712, 25594, 24440, 35069, 32088
2008 (16)	45186	From 16 keep 25817, 3744, 34571, 4709, 1676, 36378, 1004, 14172, 3352, 43081
TOTAL	194371	

S3_MED: (<i>safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.</i>)		
ANNÉE -	Syntaxe	
(NUMÉRO DE LA REQUÊTE)		
1999 (14)	11872	From 14 keep 3815, 9681, 4224, 11407, 1084, 8741, 9278, 9147, 8457, 6659
2000 (15)	13194	From 15 keep 11263, 2285, 205, 6749, 7283, 6136, 818, 2079, 10647, 9973
2001 (16)	14662	From 16 keep 1980, 7928, 11249, 62, 10348, 12885, 13362, 5161, 14629, 1698
2002 (17)	16354	From 17 keep 15762, 4792, 14986, 5053, 7623, 2401, 9225, 8196, 8758, 14732
2003 (18)	18436	From 18 keep 6329, 13864, 10072, 7558, 9383, 8094, 12451, 5573, 7420, 3937
2004 (19)	20474	From 19 keep 2612, 7498, 1949, 9840, 8365, 16295, 3629, 10646, 3603, 20118
2005 (20)	22306	From 20 keep 4487, 1522, 6395, 12523, 8825, 18289, 11009, 11057, 2506, 9037
2006 (21)	23892	From 21 keep 1893, 14439, 2130, 1647, 7725, 3224, 17929, 21837, 10611, 4089
2007 (22)	25818	From 22 keep 4788, 1780, 7356, 5034, 3705, 23281, 9219, 10766, 6576, 7545
2008 (23)	27363	From 23 keep 12111, 14665, 22220, 21992, 4591, 6710, 8656, 6976, 22891, 20235
TOTAL	194371	

SPI_MED: (<i>safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.</i>) AND (<i>exp *Risk Management/ OR exp *Medical Errors/ OR *Safety Management/</i>)		
ANNÉE -	Syntaxe	
(NUMÉRO DE LA REQUÊTE)		
1999 (12)	448	From 12 keep 301, 127, 310, 120, 32, 280, 277, 409, 70, 8
2000 (13)	825	From 13 keep 16, 26, 232, 85, 749, 578, 108, 38, 437, 338
2001 (14)	988	From 14 keep 843, 380, 128, 746, 898, 421, 4, 322, 914, 891
2002 (15)	977	From 15 keep 116, 819, 240, 677, 119, 431, 481, 725, 304, 665
2003 (16)	1281	From 16 keep 3, 317, 1018, 1169, 190, 242, 1278, 702, 746, 1270
2004 (17)	1547	From 17 keep 315, 570, 1363, 991, 1186, 396, 320, 573, 1339, 516
2005 (18)	1660	From 18 keep 986, 1163, 647, 1210, 1588, 1407, 595, 1493, 1394, 513
2006 (19)	1776	From 19 keep 1221, 1002, 1574, 1460, 436, 180, 417, 1683, 356, 910
2007 (20)	1848	From 20 keep 83, 268, 1410, 1790, 933, 1400, 622, 1595, 798, 514
2008 (21)	1884	From 21 keep 864, 598, 1357, 916, 1041, 189, 165, 1807, 859, 1150
TOTAL	13234	

SP2_MED: (<i>exp *Medical Errors/ OR *Safety Management/</i>) AND (<i>safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.</i>)		
ANNÉE -	Syntaxe	
(NUMÉRO DE LA REQUÊTE)		
1999 (11)	369	From 11 keep 24, 251, 109, 280, 239, 227, 9, 198, 3, 121
2000 (12)	718	From 12 keep 19, 182, 2, 277, 616, 515, 712, 465, 600, 591
2001 (13)	898	From 13 keep 165, 182, 90, 82, 235, 183, 128, 387, 482, 340
2002 (14)	845	From 14 keep 178, 789, 665, 105, 281, 684, 843, 40, 236, 690
2003 (15)	1114	From 15 keep 886, 121, 60, 956, 212, 863, 761, 128, 774, 319
2004 (16)	1275	From 16 keep 1120, 1104, 228, 1154, 1246, 149, 459, 190, 690, 132
2005 (17)	1401	From 17 keep 838, 103, 1095, 506, 896, 356, 267, 427, 244, 1206
2006 (18)	1493	From 18 keep 193, 730, 1271, 217, 453, 512, 1335, 434, 32, 343
2007 (19)	1560	From 19 keep 1004, 1277, 1539, 607, 109, 441, 443, 894, 1347, 489
2008 (20)	1571	From 20 keep 655, 829, 1449, 1026, 1349, 62, 326, 130, 795, 1292
TOTAL	11244	

SP3_MED:
(*safe\$.ti,ab. OR exp Safety/ OR Err\$.ti,ab. OR Adverse.ti,ab.) AND (exp *Risk Management/ OR exp Medical Errors/ OR Safety Management/ OR Medical Audit)*)

ANNÉE - (NUMÉRO DE LA REQUÊTE)	Syntaxe
1999 (13)	782 From 13 keep 548, 193, 98, 747, 641, 643, 260, 195, 1, 23
2000 (14)	1149 From 14 keep 261, 1097, 177, 893, 954, 426, 270, 1038, 1101, 927
2001 (15)	1438 From 15 keep 138, 709, 1426, 265, 1271, 176, 894, 1094, 1134, 76
2002 (16)	1467 From 16 keep 1185, 1337, 571, 374, 209, 1457, 841, 1208, 467, 1303
2003 (17)	1806 From 17 keep 1210, 1484, 338, 1691, 1412, 1227, 463, 1390, 1754, 1322
2004 (18)	2149 From 18 keep 726, 1886, 162, 2127, 1344, 86, 699, 476, 1838, 1996
2005 (19)	2316 From 19 keep 1833, 1617, 782, 611, 1647, 1665, 944, 566, 1402, 1858
2006 (20)	2575 From 20 keep 409, 847, 1966, 1267, 2564, 1855, 568, 2514, 1306, 2237
2007 (21)	2706 From 21 keep 1407, 2130, 1801, 1417, 534, 2646, 245, 956, 1020, 203
2008 (22)	2799 From 22 keep 747, 400, 764, 1460, 1477, 1808, 1031, 488, 2658, 2624
TOTAL	19187

P1_MED: **Medical Errors/*
**ANNÉE - (NUMÉRO
DE LA REQUÊTE)**

ANNÉE - (NUMÉRO DE LA REQUÊTE)	Syntaxe
1999 (3)	148 From 3 keep 59, 30, 54, 67, 80, 7, 144, 46, 136, 68
2000 (4)	440 From 4 keep 119, 419, 391, 394, 91, 132, 315, 380, 134, 418
2001 (5)	503 From 5 keep 243, 470, 168, 376, 97, 372, 311, 208, 138, 223
2002 (6)	457 From 6 keep 208, 444, 193, 6, 434, 147, 243, 194, 86, 65
2003 (7)	609 From 7 keep 392, 413, 560, 281, 218, 561, 359, 206, 158, 585
2004 (8)	607 From 8 keep 588, 525, 87, 274, 363, 330, 506, 489, 484, 403
2005 (9)	680 From 9 keep 473, 538, 599, 115, 346, 439, 82, 282, 459, 280
2006 (10)	559 From 10 keep 249, 150, 89, 481, 256, 70, 551, 42, 235, 341
2007 (11)	598 From 11 keep 399, 318, 118, 310, 35, 195, 223, 112, 275, 232
2008 (12)	586 From 12 keep 379, 102, 444, 458, 114, 405, 400, 147, 563, 224
TOTAL	5187

P2_MED:
(*exp *Medical Errors/ OR *Safety Management/ AND (patient\$.ti,ab.)*)

ANNÉE - (NUMÉRO DE LA REQUÊTE)	Syntaxe
1999 (7)	185 From 7 keep 157, 120, 102, 36, 84, 141, 24, 76, 108, 124
2000 (8)	286 From 8 keep 99, 119, 227, 201, 259, 238, 179, 121, 249, 118
2001 (9)	426 From 9 keep 202, 201, 256, 336, 286, 299, 330, 333, 338, 185
2002 (10)	502 From 10 keep 267, 475, 39, 277, 462, 184, 384, 431, 279, 444
2003 (11)	668 From 11 keep 579, 585, 326, 527, 91, 563, 637, 227, 593, 124
2004 (12)	727 From 12 keep 619, 591, 9, 265, 714, 641, 4, 373, 61, 207
2005 (13)	872 From 13 keep 53, 30, 320, 150, 51, 854, 623, 111, 545, 433
2006 (14)	946 From 14 keep 563, 16, 209, 763, 607, 458, 133, 180, 576, 415
2007 (15)	960 From 15 keep 560, 415, 598, 17, 459, 591, 86, 64, 497, 42
2008 (16)	945 From 16 keep 800, 766, 317, 23, 414, 147, 278, 49, 627, 101
TOTAL	6517

P3_MED:*(exp *Risk Management/) AND (patient\$.ti,ab.)*

ANNÉE - (NUMÉRO DE LA REQUÊTE)	Syntaxe
1999 (5)	175 From 5 keep 107, 90, 147, 115, 86, 116, 70, 12, 3, 28
2000 (6)	215 From 6 keep 132, 165, 156, 130, 166, 28, 177, 7, 212, 136
2001 (7)	339 From 7 keep 182, 239, 95, 96, 171, 74, 17, 205, 298, 301
2002 (8)	371 From 8 keep 65, 7, 312, 161, 235, 52, 328, 295, 207, 91
2003 (9)	523 From 9 keep 456, 200, 190, 467, 285, 132, 260, 393, 39, 204
2004 (10)	807 From 10 keep 172, 591, 193, 717, 55, 300, 330, 66, 246, 763
2005 (11)	1219 From 11 keep 552, 864, 68, 148, 554, 1085, 684, 541, 1198, 711
2006 (12)	1249 From 12 keep 138, 539, 721, 1211, 182, 782, 818, 979, 459, 1033
2007 (13)	1091 From 13 keep 466, 609, 233, 934, 915, 971, 110, 351, 252, 932
2008 (14)	1183 From 14 keep 781, 867, 706, 912, 701, 621, 831, 660, 834, 154
TOTAL	7172

EMBASE

S1_EMB:

(Exp Safety/ OR Safe\$.ti,ab. OR Err\$.ti,ab. OR Adverse.ti,ab) AND [(Exp Health Care Quality/ OR Risk assessment/) OR (exp *Safety/ AND *Medical Practice)]

ANNÉE -

Syntaxe

(NUMÉRO DE LA

REQUÊTE)

1999 (7)	5941	From 7 keep 3004, 2806, 1492, 4579, 2321, 220, 5658, 4927, 5611,3871
2000 (8)	7474	From 8 keep 4817, 101, 6377, 6448, 702, 5490, 1685, 3469, 3084, 158
2001 (9)	8983	From 9 keep 1353, 7914, 1681, 7152, 1696, 4241, 6975, 3080, 6939, 7829
2002 (10)	11249	From 10 keep 1500, 6425, 4564, 1988, 4463, 11051, 1828, 4094, 8213, 2712
2003 (11)	13263	From 11 keep 11305, 3775, 8955, 10177, 9755, 996, 6399, 2759, 8078, 225
2004 (12)	14304	From 12 keep 8219, 12414, 7168, 4817, 6755, 1702, 9194, 11396, 9656, 1730
2005 (13)	15866	From 13 keep 4927, 9194, 15486, 14766, 10440, 9954, 6528, 12301, 1723, 12236
2006 (14)	17729	From 14 keep 6255, 6675, 8660, 2642, 11835, 8010, 15764, 8793, 13896, 4312
2007 (15)	18589	From 15 keep 2374, 11905, 655, 15662, 675, 406, 15886, 10722, 9022, 12498
2008 (16)	19358	From 16 keep 11979, 13120, 2905, 9586, 16685, 5468, 9875, 17759, 5857, 5763
TOTAL	132756	

S2_EMB:

(Exp Safety/ OR Safe\$.ti,ab. OR Err\$.ti,ab. OR Adverse.ti,ab) AND [(exp Health Care Quality/ OR ((Exp Safety/ OR Err\$.ti,ab) AND (Exp Health Care Organization/ OR Exp Health Care/ OR Exp Health Care System/ OR Hospital\$.ti,ab.)]

ANNÉE -

Syntaxe

(NUMÉRO DE LA

REQUÊTE)

1999 (9)	7382	From 9 keep 3134, 293, 3579, 6293, 408, 7243, 2326, 929, 3999, 6322
2000 (10)	9306	From 10 keep 9229, 6099, 2910, 6125, 1395, 72, 6155, 6283, 8237, 5886
2001 (11)	11351	From 11 keep 2165, 8260, 6713, 3976, 7011, 10159, 8601, 1595, 250, 11086
2002 (12)	13809	From 12 keep 849, 10253, 6141, 4756, 12560, 11778, 12979, 6704, 12888, 6955
2003 (13)	16018	From 13 keep 7317, 4044, 375, 10341, 2032, 2905, 11368, 7791, 4441, 1674
2004 (14)	16843	From 14 keep 12060, 6199, 12873, 981, 4731, 14795, 2059, 161, 6504, 1087
2005 (15)	18571	From 15 keep 7144, 11137, 14622, 1904, 16422, 10481, 10352, 12693, 17277, 18005
2006 (16)	20746	From 16 keep 3914, 19269, 17005, 13948, 1764, 13068, 14971, 1206, 11875, 10694
2007 (17)	21730	From 17 keep 8424, 20679, 7495, 3477, 20119, 11800, 11473, 6471, 1938, 5346
2008 (18)	22377	From 18 keep 8582, 4450, 10030, 6771, 3628, 5030, 8225, 16029, 16435, 17018
TOTAL	158133	

S3_EMB:

(exp Medical Error/ **OR** Patient Safety/ **OR** Exp Iatrogenic disease/) **OR** [(Exp Safety/ **OR** Safe\$.ti.ab. **OR** Err\$.ti.ab. **OR** Adverse.ti.ab.) **AND** [(exp Health Care Quality/) **OR** (exp *Safety/ **AND** (Exp Health Care Organization/ **OR** Exp Health Care/ **OR** Medic\$.ti.ab. **OR** Exp Health Care System/ **OR** Hospital\$.ti.ab.)]]

ANNÉE -

Syntaxe

(NUMÉRO DE LA
REQUÊTE)

1999 (11)	11271	From 11 keep 4258, 6066, 4856, 1062, 5505, 4146, 1933, 6115, 1477, 3444
2000 (12)	14150	From 12 keep 10954, 4277, 10224, 892, 5145, 9398, 12208, 11272, 11113, 8406
2001 (13)	18474	From 13 keep 16723, 4270, 16155, 14958, 14464, 416, 8339, 4505, 2477, 5005
2002 (14)	22931	From 14 keep 7510, 7745, 13747, 4333, 10732, 3258, 18036, 21567, 5173, 1468
2003 (15)	30399	From 15 keep 24469, 14546, 26668, 21533, 7044, 30333, 13029, 13001, 28985, 10909
2004 (16)	33685	From 16 keep 12594, 33625, 26987, 8179, 10464, 16737, 23806, 2719, 12416, 29692
2005 (17)	35087	From 17 keep 25782, 21106, 21297, 20373, 19646, 5900, 23869, 2801, 22176, 21084
2006 (18)	40652	From 18 keep 36479, 38371, 40115, 12173, 26406, 12052, 14230, 25051, 29616, 28194
2007 (19)	43724	From 19 keep 43102, 3218, 19599, 21863, 1356, 28671, 43328, 959, 39516, 27183
2008 (20)	45400	From 20 keep 11641, 1297, 30624, 41589, 25080, 32772, 41925, 15503, 10772, 2501
TOTAL	295773	

SP1_EMB:

[(Exp Safety/ **OR** Err\$.ti.ab.) **AND** (Health Care Quality/)] **OR** [(Exp Safety/ **OR** Safe\$.ti.ab. **OR** Err\$.ti.ab. **OR** Adverse.ti.ab.) **AND** (Exp Health Care Organization/ **OR** Health\$.ti.ab.)]

ANNÉE -

Syntaxe

(NUMÉRO DE LA
REQUÊTE)

1999 (9)	4496	From 9 keep 1442, 4257, 1203, 1644, 212, 1605, 4322, 218, 4462, 1781
2000 (10)	5336	From 10 keep 1895, 413, 3651, 4586, 4488, 1436, 2521, 5110, 3734, 426
2001 (11)	5963	From 11 keep 3747, 913, 3317, 5376, 4323, 5859, 5549, 2184, 1036, 4004
2002 (12)	7320	From 12 keep 23, 2136, 2122, 164, 2459, 4366, 773, 374, 6528, 2538
2003 (13)	8421	From 13 keep 6182, 2325, 1496, 2080, 1809, 1867, 6309, 1690, 4455, 6290
2004 (14)	9517	From 14 keep 6083, 3736, 7290, 5892, 2766, 226, 8104, 1561, 7317, 5115
2005 (15)	10499	From 15 keep 7407, 1405, 9690, 999, 4568, 2666, 6123, 9303, 623, 9206
2006 (16)	11843	From 16 keep 1132, 5443, 6954, 9262, 6406, 7098, 7554, 8756, 10917, 1454
2007 (17)	12485	From 17 keep 4299, 9686, 4077, 641, 10354, 9634, 12349, 5847, 52, 3454
2008 (18)	13470	From 18 keep 12028, 4173, 7851, 2188, 9404, 9621, 11708, 5328, 11492, 6693
TOTAL	89350	

SP2_EMB:

(Exp Safety/ **OR** Safe\$.ti.ab. **OR** Err\$.ti.ab. **OR** Adverse.ti.ab.) **AND** [(Health Care Quality/ **OR** ((Exp Safety/ **OR** Err\$.ti.ab.) **AND** (Exp Health Care Organization/ **OR** Health\$.ti.ab.)]]

ANNÉE -

Syntaxe

(NUMÉRO DE LA
REQUÊTE)

1999 (9)	1683	From 9 keep 432, 1012, 663, 1476, 21, 1335, 1562, 615, 520, 297
2000 (10)	2291	From 10 keep 2232, 1019, 798, 1092, 1975, 1256, 2096, 521, 987, 1787
2001 (11)	2829	From 11 keep 1298, 613, 1178, 333, 1860, 1723, 528, 537, 550, 65
2002 (12)	3473	From 12 keep 3385, 2098, 2186, 622, 745, 2570, 720, 314, 1247, 271
2003 (13)	4176	From 13 keep 228, 296, 3775, 2550, 2342, 974, 2780, 1793, 2273, 3391
2004 (14)	4759	From 14 keep 3433, 2262, 895, 3795, 2784, 3781, 1104, 4007, 2833, 864
2005 (15)	5113	From 15 keep 2673, 4391, 3349, 2615, 3822, 2586, 4977, 626, 4427, 2391
2006 (16)	5839	From 16 keep 2848, 1967, 5091, 1213, 635, 5037, 366, 1223, 3527, 440
2007 (17)	5869	From 17 keep 41, 4456, 2621, 5630, 3303, 2219, 1850, 1174, 5124, 2047
2008 (18)	6128	From 18 keep 4603, 3554, 1651, 1874, 1262, 2031, 1158, 5903, 1714, 324
TOTAL	42160	

SP3_EMB:(Exp Safety/ **OR** Safe\$.ti.ab. **OR** Err\$.ti.ab. **OR** Adverse.ti.ab.) **AND** [Health Care Quality **OR** (Exp Health Care Organization/ **OR** Health\$.ti.ab.)]**ANNÉE -
(NUMÉRO DE LA
REQUÊTE)****Syntaxe**

1999 (7)	4564	From 7 keep 4316, 1624, 2194, 3755, 4381, 798, 860, 683, 1080, 1700
2000 (8)	5399	From 8 keep 4231, 353, 814, 4154, 4665, 4521, 2399, 2519, 1509, 1980
2001 (9)	6045	From 9 keep 247, 3782, 2627, 4172, 3212, 4504, 3019, 1263, 3431, 5725
2002 (10)	7389	From 10 keep 24, 6120, 165, 6519, 5288, 291, 3388, 235, 7021, 6630
2003 (11)	8500	From 11 keep 2674, 1345, 5062, 1479, 6492, 5531, 2535, 4243, 4409, 3226
2004 (12)	9622	From 12 keep 1681, 2988, 2651, 6674, 1867, 7519, 1885, 7617, 5288, 1988
2005 (13)	10587	From 13 keep 9538, 2711, 423, 6024, 2235, 9631, 8515, 3151, 8931, 3198
2006 (14)	11934	From 14 keep 6431, 3546, 4556, 9495, 8810, 11026, 6991, 2483, 39, 300
2007 (15)	12589	From 15 keep 10878, 944, 3320, 8011, 9369, 11469, 6436, 4436, 4346, 1463
2008 (16)	13582	From 16 keep 9407, 10783, 3438, 5175, 6443, 6195, 148, 5534, 1871, 5407
TOTAL	90211	

PI_EMB:(Exp Safety/ **OR** Err\$.ti.ab.) **AND** (patient\$.ti.ab.) **AND** (Exp Health Care Organization/ **OR** Exp Health Care/ **OR** Medic\$.ti.ab. **OR** Exp Health Care System/ **OR** Hospital\$.ti.ab.)**ANNÉE -
(NUMÉRO DE LA
REQUÊTE)****Syntaxe**

1999 (6)	1111	From 6 keep 737, 781, 793, 669, 326, 594, 123, 955, 135, 53
2000 (7)	1452	From 7 keep 825, 92, 827, 904, 440, 463, 495, 529, 947, 1140
2001 (8)	2152	From 8 keep 1452, 1767, 1638, 1277, 1989, 2061, 676, 89, 2106, 303
2002 (9)	2682	From 9 keep 1118, 293, 1606, 1488, 1353, 1466, 2589, 599, 1508, 226
2003 (10)	3386	From 10 keep 676, 3277, 1869, 3231, 3229, 2647, 1757, 23, 536, 2149
2004 (11)	3441	From 11 keep 2352, 3037, 2679, 306, 315, 693, 3242, 1766, 2965, 3246
2005 (12)	3842	From 12 keep 2040, 2991, 3013, 2222, 151, 1805, 680, 3583, 836, 377
2006 (13)	4484	From 13 keep 3659, 2283, 3278, 3733, 112, 51, 3185, 1540, 3886, 3923
2007 (14)	4444	From 14 keep 792, 865, 1500, 3948, 3043, 723, 1685, 1147, 1171, 3907
2008 (15)	4885	From 15 keep 4175, 4177, 775, 3673, 1729, 3956, 3953, 2490, 889, 1058
TOTAL	31879	

P2_EMB:(Exp Safety/ **OR** Err\$.ti.ab.) **AND** [(Patient Care/) **OR** [(patient\$.ti.ab.) **AND** (Exp Health Care Organization/ **OR** Exp Health Care/ **OR** Medic\$.ti.ab. **OR** Exp Health Care System/ **OR** Hospital\$.ti.ab.)]]**ANNÉE -
(NUMÉRO DE LA
REQUÊTE)****Syntaxe**

1999 (10)	1117	From 10 keep 1016, 224, 293, 958, 485, 238, 556, 1071, 295, 394
2000 (11)	1465	From 11 keep 534, 1007, 204, 173, 35, 226, 1327, 372, 1253, 942
2001 (12)	2169	From 12 keep 1360, 1292, 392, 1676, 1791, 1733, 361, 714, 1669, 1045
2002 (13)	2690	From 13 keep 2467, 322, 1301, 85, 697, 2537, 1128, 2552, 118, 1117
2003 (14)	3405	From 14 keep 6, 81, 2012, 2357, 838, 1811, 2622, 2194, 3181, 2118
2004 (15)	3458	From 15 keep 2435, 2224, 3038, 2118, 2102, 1907, 2134, 1252, 400, 1695
2005 (16)	3868	From 16 keep 3683, 1448, 1585, 1116, 2769, 1859, 112, 2555, 1567, 3545
2006 (17)	4516	From 17 keep 2787, 3955, 379, 2999, 1336, 689, 555, 2217, 261, 2003
2007 (18)	4489	From 18 keep 1845, 3698, 1291, 4239, 1004, 1587, 2713, 493, 152, 44
2008 (19)	4935	From 19 keep 884, 2313, 4741, 869, 1559, 1220, 1713, 4385, 4329, 4367
TOTAL	32112	

P3_EMB:

(Exp Safety/ **OR** Err\$.ti.ab.) **AND** [(Patient Care/) **OR** [(patient\$.ti.ab. **OR** Human/ **OR** Human.ti.ab.) **AND** (Exp Health Care Organization/ **OR** Health\$.ti.ab.)]]

ANNÉE -**Syntaxe****(NUMÉRO DE LA
REQUÊTE)**

1999 (9)	1051	From 9 keep 927, 68, 429, 949, 699, 1029, 824, 220, 64, 775
2000 (10)	1535	From 10 keep 1378, 1028, 1511, 1532, 1520, 1229, 3, 345, 682, 900
2001 (11)	1970	From 11 keep 703, 981, 503, 1151, 1727, 861, 1618, 1763, 1407, 1759
2002 (12)	2518	From 12 keep 1892, 194, 1084, 1977, 1593, 603, 429, 37, 597, 1063
2003 (13)	3104	From 13 keep 1244, 2102, 2443, 22, 1290, 1331, 225, 3102, 139, 1210
2004 (14)	3449	From 14 keep 538, 162, 2466, 2177, 2893, 1628, 1973, 2115, 667, 2814
2005 (15)	3803	From 15 keep 626, 691, 2649, 612, 3371, 2053, 3370, 1182, 1237, 1079
2006 (16)	4377	From 16 keep 1529, 1347, 3438, 4105, 108, 2677, 131, 2861, 315, 3313
2007 (17)	4370	From 17 keep 4206, 1238, 273, 3757, 309, 105, 2083, 1701, 2357, 965
2008 (18)	4644	From 18 keep 2745, 4283, 2634, 1025, 2648, 1719, 1749, 462, 3492, 4269
TOTAL	30821	

CINAHL¹¹**S1-CIN**

[(Tx Health OR Tx Care* OR Tx Prevent*) AND (Tx Safe* OR Tx Err*) AND (MH "Child" OR Tx Patient* or Tx Human)] OR MH "Health Care Errors+"*

ANNÉE	Total des articles repérés	Numéro des articles sélectionnés
999	4479	1851, 2129, 4288, 3174, 4215, 4361, 253, 4057, 2797, 2491
2000	5642	86; 868; 1483, 1636, 1740, 2450, 3351, 3889, 4383, 4862
2001	6992	2950, 5890, 4558, 2964, 1965, 1300, 5678, 5114, 6098, 5333
2002	9013	639, 6052, 3349, 7305, 5146, 7618, 1317, 8352, 8427, 7957
2003	11803	744, 8546, 8105, 4222, 8538, 10512, 9394, 421, 5430, 380
2004	13298	8300, 5574, 6992, 5437, 10012, 2687, 2795, 10295, 7312, 5583
2005	14796	3282, 3515, 3530, 4617, 6585, 8468, 9511, 9671, 9916, 13873
2006	20034	3871, 4036, 6783, 14914, 13432, 18412, 15315, 9916, 6233, 5838
2007	21086	3621, 16324, 4906, 2180, 12570, 17902, 18574, 4145, 4564, 1338
2008	23925	7846, 12690, 335, 7549, 8785, 21452, 18865, 2019, 6538, 19754
TOTAL	131068	

S2-CIN:

(Tx Health or Tx Care* or Tx Prevent*) AND (Tx Safe* or Tx Err*)*

ANNÉE	Total des articles repérés	Numéro des articles sélectionnés
1999	7920	6792, 4797, 715, 329, 446, 6422, 5753, 5574, 740, 6252
2000	9837	6308, 8477, 6483, 1721, 2432, 767, 3218, 7929, 4629, 4514
2001	11493	6068, 5162, 6330, 8053, 1465, 694, 7962, 284, 9650, 634
2002	14199	2212, 6568, 7959, 706, 9473, 2483, 5098, 6887, 14107, 10482
2003	17660	11728, 1270, 8833, 12572, 13394, 2242, 1980, 8369, 15854, 10338
2004	20241	8070, 5871, 16, 6107, 1571, 8181, 2103, 1240, 4193, 18983
2005	22685	13411, 17020, 22597, 1557, 60, 19624, 6242, 11366, 73, 16378
2006	29387	1728, 21870, 22678, 4902, 22888, 7614, 8673, 2144, 25106, 7063
2007	31427	6344, 5918, 13346, 2948, 31182, 10490, 19481, 22114, 24385, 667
2008	36773	14767, 19146, 13949, 34549, 36530, 19663, 30515, 19386, 7389, 14278
TOTAL	201622	

¹¹ CINAHL ne permet pas de rédiger une syntaxe pour sélectionner les articles comme c'est le cas dans Medline et Embase, il a fallu sélectionner chacun des articles manuellement sur la page de résultats pour les envoyer dans un dossier personnel avant de les exporter dans le logiciel Endnote.

S3_CIN:

[(Tx Health* OR Tx Care* OR Tx Prevent*) AND (Tx Safe* OR Tx Err*)] AND (MH "Child" OR Tx Patient* OR Tx Human).

ANNÉE	Total des articles repérés	Numéro des articles sélectionnés
1999	4478	1917, 2518, 1191, 2840, 2824, 1597, 3479, 3952, 1771, 154
2000	5635	1130, 1060, 1563, 2823, 4743, 1669, 3530, 5441, 3026, 3873
2001	6973	3081, 1516, 461, 14, 956, 2260, 251, 3163, 6431, 2643
2002	8975	878, 1746, 658, 8202, 4533, 7109, 8685, 6297, 5221, 3828
2003	11755	10806, 9846, 4054, 5963, 1568, 8130, 9783, 5641, 4480, 5581
2004	13259	12701, 7768, 2638, 1896, 12286, 3393, 12818, 110, 4006, 1303
2005	14748	7317, 3597, 1154, 10443, 6915, 160, 63, 161, 4765, 5757
2006	19997	13384, 6796, 8652, 15394, 4592, 10255, 16325, 4955, 3860, 3990
2007	21033	5511, 16189, 9948, 134, 9555, 8259, 13397, 20383, 12481, 16668
2008	23845	1396, 1241, 1776, 19414, 22935, 14460, 7243, 20764, 303, 19656
TOTAL	130698	

SP1_CIN: MH "Health Care Errors+"

ANNÉE	Total des articles repérés	Numéro des articles sélectionnés
1999	492	373, 466, 197, 375, 398, 35, 68, 62, 337, 259
2000	936	239, 156, 918, 858, 519, 351, 244, 241, 83, 151
2001	971	694, 472, 888, 188, 553, 23, 406, 395, 866, 88
2002	1082	403, 1068, 606, 187, 180, 155, 990, 1012, 876, 1019
2003	1411	907, 621, 382, 72, 110, 27, 1072, 1066, 757, 209
2004	1448	491, 798, 938, 483, 525, 689, 1215, 1030, 896, 45
2005	1669	175, 825, 1492, 1039, 1362, 1343, 1073, 347, 829, 398
2006	1968	963, 1329, 1321, 263, 1456, 1566, 1186, 171, 288, 404
2007	1974	345, 541, 1885, 498, 1305, 1036, 735, 75, 1358, 1097
2008	2453	461, 1749, 1054, 871, 949, 1878, 118, 1409, 975, 1973
TOTAL	14404	

SP2_CIN: [(Tx Err* AND (Tx Health* OR Tx Care* OR Tx Prevent*)] OR MH "Risk Management+") AND (MH "Child" OR Tx Patient* OR Tx Human)] OR MH "Health Care Errors+".

ANNÉE	Total des articles repérés	Numéro des articles sélectionnés
1999	2537	1856, 1151, 804, 844, 123, 1097, 419, 2528, 565, 336
2000	3476	457, 2417, 908, 2321, 749, 1760, 1939, 2731, 2839, 3379
2001	3897	808, 3182, 2514, 3694, 2480, 2074, 1625, 1232, 3042, 252
2002	4615	187, 3814, 3647, 742, 3359, 2566, 3605, 817, 1435, 3251
2003	6088	3084, 5516, 2247, 1432, 5797, 5110, 3384, 4068, 1974, 3721
2004	6770	5821, 4763, 3147, 2582, 5288, 4867, 5890, 931, 6442, 1051
2005	7267	4506, 2467, 5058, 5343, 2548, 2745, 6251, 6658, 5846, 4966
2006	9586	1264, 7224, 2209, 2762, 6791, 7275, 2301, 6772, 588, 516
2007	9990	9676, 4064, 5513, 4788, 4107, 8808, 4393, 5502, 5014, 6210
2008	12287	10637, 9359, 3359, 11045, 11891, 1164, 6120, 11837, 4525, 1100
TOTAL	66513	

SP3_CIN: [([Tx Err* AND (Tx Health* OR Tx Care* OR Tx Prevent*)] OR MH "Risk Management+") AND (MH "Child" OR Tx Patient* OR Tx Human)]

ANNÉE	Total des articles repérés	Numéro des articles sélectionnés
1999	2200	1834, 1461, 290, 2066, 1782, 132, 1559, 382, 1940, 291
2000	2913	603, 1957, 2228, 1444, 1593, 765, 2138, 231, 296, 405
2001	3414	1794, 3204, 3123, 2002, 3322, 1755, 1556, 1436, 309, 779
2002	4093	2011, 1548, 1193, 1415, 3500, 889, 1282, 1155, 3632, 926
2003	5486	3355, 1073, 2105, 505, 5311, 4948, 4431, 279, 3185, 3002
2004	6271	1517, 3080, 4630, 3991, 2963, 4167, 5194, 2813, 2346, 1437
2005	6628	3468, 2973, 5591, 677, 4783, 2768, 1800, 2694, 441, 1423
2006	8918	3458, 2061, 3942, 1838, 5391, 191, 4981, 8899, 4325, 4653
2007	9250	464, 3063, 1444, 2374, 4330, 84, 3849, 1605, 8798, 746
2008	11337	9980, 1593, 5074, 8189, 6026, 2984, 5765, 3568, 360, 3878
TOTAL	60510	

P1_CIN: Tx Harm* AND Tx Event*

ANNÉE	Total des articles repérés	Numéro des articles sélectionnés
1999	786	684, 517, 282, 71, 153, 19, 583, 60, 242, 464
2000	918	687, 422, 646, 115, 87, 280, 671, 16, 663, 11
2001	985	68, 145, 457, 225, 151, 514, 907, 587, 152, 933
2002	1396	749, 789, 223, 593, 1193, 914, 516, 587, 750, 165
2003	1672	463, 1076, 51, 812, 329, 207, 772, 83, 1116, 1432
2004	1963	1373, 1086, 943, 1320, 933, 368, 584, 1717, 1360, 683
2005	2244	1301, 1374, 392, 611, 1002, 1390, 2102, 645, 259, 209
2006	3226	1045, 271, 2640, 487, 1596, 644, 2625, 2790, 1413, 115
2007	3400	471, 1010, 1751, 807, 1277, 3214, 3201, 1269, 416, 1617
2008	4023	834, 470, 1703, 1632, 3523, 3485, 312, 1028, 2361, 514
TOTAL	20613	

P2_CIN:
Tx Err* AND Tx Event*

ANNÉE -	Total des articles repérés	Numéro des articles sélectionnés
1999	807	183, 585, 375, 71, 497, 441, 95, 408, 184, 302
2000	1009	430, 573, 152, 345, 23, 932, 631, 329, 4, 930
2001	1162	845, 687, 536, 596, 1158, 586, 151, 945, 190, 902
2002	1537	383, 459, 81, 240, 369, 372, 1219, 14, 902, 547
2003	1907	381, 241, 1327, 592, 1777, 1513, 409, 1072, 1593, 64
2004	2224	1157, 1717, 2203, 1924, 30, 318, 740, 340, 1213, 1285
2005	2321	1492, 134, 1585, 2074, 2146, 483, 1420, 2159, 780, 1153
2006	3493	2475, 3368, 2867, 2231, 3323, 2638, 1508, 1318, 2251, 3291
2007	3808	2629, 2843, 2391, 1770, 2909, 692, 884, 2615, 3002, 2258
2008	4370	268, 2846, 28, 3859, 1575, 4189, 255, 2263, 2999, 3622
TOTAL	22638	

P3_CIN:

[(Tx Harm AND Tx Event*) OR MH "Risk Management+"] OR (MH "Child" OR Tx Patient* OR Tx Human)*

ANNÉE -	Total des articles repérés	Numéro des articles sélectionnés
1999	918	444, 336, 911, 536, 375, 104, 20, 574, 572, 730
2000	1035	142, 38, 286, 363, 539, 752, 542, 282, 534, 311
2001	1182	919, 885, 411, 1111, 126, 539, 904, 146, 532, 121
2002	1529	91, 225, 52, 1278, 1267, 555, 1431, 1233, 1254, 891
2003	1895	1736, 1328, 1223, 577, 1527, 574, 1768, 1325, 6, 1230
2004	2177	1857, 159, 666, 1312, 1711, 1682, 821, 212, 972, 125
2005	2477	1729, 2472, 1945, 1768, 1975, 789, 678, 821, 1324, 2257
2006	3461	2996, 3238, 3260, 2176, 2698, 317, 1912, 935, 1991, 3309
2007	3671	3268, 1015, 444, 1515, 2484, 160, 1753, 3344, 552, 2976
2008	4362	2017, 60, 272, 3999, 1285, 954, 157, 3960, 383, 1978
TOTAL	22707	

ⁱ www.npsf.org

ⁱⁱ <http://www.ulrichsweb.com/ulrichsweb/>

ⁱⁱⁱ For each of the 55 journals, and for the years 2000, 2002, 2004 and 2006, a quick search was performed in MEDLINE using the strategy (Medical Errors/ or (Safety/ and exp Patient/)). The records thus found were called "patient safety records." The journals were then sorted according to the number of patient safety records.

^{iv} <http://www.trialstat.com/>

^v A software tool for publishing and managing bibliographies. www.endnote.com.