

Directives pour l'évaluation des études réalisées à partir de données de santé collectées en routine : RECORD (REporting of studies Conducted using Observational Routinely-collected health Data)

Journal:	CMAJ
Manuscript ID	CMAJ-18-1309
Manuscript Type:	Research - Other
Date Submitted by the Author:	02-Oct-2018
Complete List of Authors:	Benchimol, Eric; Children's Hospital of Eastern Ontario, Gastroenterology, Hepatology and Nutrition Smeeth, L; London School of Hygiene and Tropical Medicine Guttman, Astrid; Institute for Clinical Evaluative Sciences, Harron, Katie; University College London Moher, David; Ottawa Hospital Research Institute, Clinical Epidemiology Program Petersen, Irene; Uiversity College London, Research Department of Primary Care and Population Health Sorensen, Henrik; Aarhus University, Clinical Epidemiology Januel, Jean-Marie; L'Université Sorbonne Paris Cité, l'Institut du Management Von Elm, Erik; Université de Lausanne, Institut de médecine sociale et préventive Langan, Sinéad; London School of Hygiene and Tropical Medicine
Keywords:	Statistics and Research Methods, Epidemiology, Evidence-based Medicine, Medical Informatics
More Detailed Keywords:	routinely collected health data, health administrative data, reporting guidelines
Abstract:	Les données de santé collectées en routine et obtenues à des fins administratives et cliniques sans objectifs de recherche spécifiques a priori, sont de plus en plus utilisées pour la recherche. L'évolution rapide et la disponibilité de ces données ont révélé des problèmes qui ne sont pas abordés dans les outils d'évaluation existantes, tel que STROBE (Strengthening the Reporting of Observational Studies in Epidemiology). RECORD (REporting of studies Conducted using Observational Routinely collected health Data) a été développé pour combler ces lacunes. RECORD représente une extension de STROBE pour évaluer des éléments spécifiques aux études observationnelles en utilisant des données de santé collectées en routine. RECORD propose une liste de 13 éléments à évaluer concernant le titre, le résumé, l'introduction, les méthodes, les résultats et la section discussion des articles, ainsi que d'autres informations requises pour établir la qualité des études rapportées dans ces articles. Cet article

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	<p>méthodologique présente la liste des éléments d'évaluation ainsi que des informations explicatives pour réaliser et interpréter les évaluations. Des exemples sont également proposés pour chaque élément de la liste d'évaluation. Cet article présente une version en Français de l'outil d'évaluation RECORD. Accompagné de son site Internet (http://www.record-statement.org) son objectif est une large diffusion de l'outil d'évaluation RECORD pour faciliter sa mise en œuvre. Grâce à RECORD, les auteurs, les éditeurs de revues et les évaluateurs d'articles (c.-à-d., évaluation par les pairs) peuvent promouvoir la transparence des articles rapportant des études observationnelles originales à partir des données de santé collectées en routine.</p>

SCHOLARONE™
Manuscripts

Confidential

Directives pour l'évaluation des études réalisées à partir de données de santé collectées en routine : RECORD (REporting of studies Conducted using Observational Routinely-collected health Data)

Eric I. Benchimol^{1,2*}, Liam Smeeth³, Astrid Guttmann^{2,4}, Katie Harron³, David Moher⁵, Irene Petersen⁶, Henrik T. Sørensen⁷, Jean-Marie Januel⁸, Erik von Elm^{9‡}, Sinéad M. Langan^{3‡*},
Comité de travail RECORD ¶

Traduit par :

Jean-Paul Salameh (L'hôpital d'Ottawa Institut de Recherche, Ottawa, ON, Canada, et l'École d'Épidémiologie et de Santé Publique, Université d'Ottawa, Ottawa, ON, Canada).
Yara Boutros (École de Traducteurs et d'Interprètes de Beyrouth, Université Saint-Joseph de Beyrouth, Liban)

1. Institut de recherche du Centre hospitalier pour enfants de l'est de l'Ontario, Département de pédiatrie et École d'épidémiologie et de sante publique, Université d'Ottawa, Ottawa, Canada
2. Institute for Clinical Evaluative Sciences, Toronto, Canada
3. London School of Hygiene and Tropical Medicine, Londres, Royaume-Uni
4. Hospital for Sick Children, Department of Paediatrics and Institute of Health Policy, Management and Evaluation, Université de Toronto, Toronto, Canada
5. L'hôpital d'Ottawa, Ottawa, Canada, et École d'épidémiologie et de sante publique, Université d'Ottawa, Ottawa, Canada
6. Department of Primary Care and Public Health, University College London, Londres, Royaume-Uni
7. Department of Clinical Epidemiology, Aarhus University, Aarhus, Danemark
8. Management des organisations de santé, l'Institut du Management, L'Université Sorbonne Paris Cité, Paris, France.
9. Cochrane Suisse, Institut de médecine sociale et préventive, Université de Lausanne, Lausanne, Suisse

‡ Les auteurs principaux ont contribué de manière égale à ce travail.

¶ La composition du comité de travail RECORD est indiquée dans les remerciements.

* ebenchimol@cheo.on.ca (BEI) ; sinead.langan@lshtm.ac.uk (SML)

Financement : Instituts de recherche en santé du Canada (numéro de subvention 130512), Fonds national suisse de la recherche scientifique (numéro de subvention IZ32Z0_147388 / 1) et Département d'épidémiologie clinique de l'Université d'Aarhus. Les bailleurs de fonds n'ont joué aucun rôle dans la conception de l'étude, la collecte et l'analyse des données, la décision de publier ou la préparation du manuscrit.

Intérêts concurrents : DM siège au comité de rédaction de PLOS Medicine. Les autres auteurs n'ont aucun conflit d'intérêt à déclarer.

Abréviations: CPRD, Clinical Practice Research Datalink; GPRD, base de données de recherche en médecine générale; HSMR, hospital standardised mortality ratio; CIM, Classification internationale des maladies; ISC, New South Wales Inpatient Statistics Collection; MDC, New South Wales Midwives Data Collection; MeSH, Medical Subject Heading; mHealth apps, applications de santé mobile; NHS, National Health Service; NSCLC, non-small cell lung cancer; PET, positron emission tomography; PICANet, Paediatric Intensive Care Audit Network; RECORD, REporting of studies Conducted using Observational Routinely collected health Data; SEER, Surveillance, Epidemiology and End Results; SNIIRAM, Système national d'information inter-régimes de l'assurance maladie; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

Provenance : non commandée ; examen externe par des pairs

Résumé

Les données de santé collectées en routine et obtenues à des fins administratives et cliniques sans objectifs de recherche spécifiques a priori, sont de plus en plus utilisées pour la recherche. L'évolution rapide et la disponibilité de ces données ont révélé des problèmes qui ne sont pas abordés dans les outils d'évaluation existantes, tel que STROBE (Strengthening the Reporting of Observational Studies in Epidemiology). RECORD (REporting of studies Conducted using Observational Routinely collected health Data) a été développé pour combler ces lacunes. RECORD représente une extension de STROBE pour évaluer des éléments spécifiques aux études observationnelles en utilisant des données de santé collectées en routine. RECORD propose une liste de 13 éléments à évaluer concernant le titre, le résumé, l'introduction, les méthodes, les résultats et la section discussion des articles, ainsi que d'autres informations requises pour établir la qualité des études rapportées dans ces articles. Cet article méthodologique présente la liste des éléments d'évaluation ainsi que des informations explicatives pour réaliser et interpréter les évaluations. Des exemples sont également proposés pour chaque élément de la liste d'évaluation. Cet article présente une version en Français de l'outil d'évaluation RECORD. Accompagné de son site Internet (<http://www.record-statement.org>) son objectif est une large diffusion de l'outil d'évaluation RECORD pour faciliter sa mise en œuvre. Grâce à RECORD, les auteurs, les éditeurs de revues et les évaluateurs d'articles (c.-à-d., évaluation par les pairs) peuvent promouvoir la transparence des articles rapportant des études observationnelles originales à partir des données de santé collectées en routine.

Introduction

La disponibilité croissante des données générées par les prestataires de soins de santé ainsi que par la surveillance organisée de l'incidence des maladies et des indicateurs de l'évolution des maladies a transformé le caractère de la recherche en santé. Les données de santé collectées en routine sont définies comme étant des données collectées sans que des questions de recherche spécifiques ne soient développées de manière a priori [1]. Ces données représentent pourtant de nombreuses ressources pour la recherche (ex., registres des maladies), la gestion

1
2
3 clinique (ex. bases de données de soins primaires), la planification du système de santé (ex.,
4 données médico-administratives), la documentation clinique des soins (ex., dossiers patients
5 électroniques), ou la surveillance épidémiologique (ex., registres des cancers et données des
6 rapports de santé publique). Ces données, recueillies dans divers contextes de soins de santé et de
7 lieux géographiques, offrent des opportunités pour développer des recherches innovantes,
8 efficaces et rentables pour éclairer les décisions cliniques en médecine, pour la planification des
9 services de santé et en santé publique [2]. À l'échelle internationale, les gouvernements et les
10 instances de financement de la santé ont priorisé l'utilisation de ces données pour améliorer les
11 soins offerts aux patients, transformer la recherche en santé et augmenter l'efficacité des soins
12 [3].

13
14
15 Malgré le fait que l'explosion de la disponibilité de ces données puisse offrir des
16 opportunités significatives pour répondre à des questions de recherche urgentes, elles posent
17 cependant des défis à ceux qui entreprennent et évaluent la recherche et mettent en œuvre des
18 décisions issues des conclusions à partir d'études utilisant ces données. Les très nombreuses
19 sources de données de santé collectées en routine et l'expansion rapide du champ de ces données
20 rendent difficile l'identification de leurs avantages et de leurs limites, mais aussi des biais
21 associés aux sources de données individuelles. Un rapport incomplet ou inadéquat des études
22 observationnelles basée sur des données de routine en santé exacerbe les défis liés à l'utilisation
23 de ces données. Une analyse systématique d'un échantillon d'études utilisant des sources de
24 données collectées de manière routinière a identifié de nombreux domaines où le rapport est
25 incomplet ou peu clair [4]. Les lacunes comprennent des informations inadéquates ou
26 manquantes concernant le codage des facteurs d'expositions et des résultats, ainsi que dans les
27 détails concernant la concordance des informations entre deux sources de données différentes.
28 Deux revues systématiques ont montré également des résultats médiocres concernant la
29 validation des données de routine [5,6], qui peuvent masquer différents types de biais, entraver
30 les efforts pour réaliser des méta-analyses à partir de ces données, et conduire à des conclusions
31 erronées.

32
33
34 Des directives destinées à évaluer les études ont été élaborées pour guider un processus
35 qui couvre un large éventail de points clés allant de la conception au contexte de telles études,
36 afin d'établir la qualité de ces dernières [8,9]. En particulier, STROBE (Strengthening the
37 Reporting of Observational Studies in Epidemiology) qui a été élaborée pour améliorer la
38 transparence dans les études observationnelles [10,11] et a été largement adoptée et approuvée
39 par la plupart des revues scientifiques médicales. Il a été démontré que la mise en œuvre de
40 STROBE dans le processus éditorial améliorerait la qualité des études publiées [12, 13]. La plupart
41 des études conduites à partir de données collectées en routine répondent aux caractéristiques des
42 études observationnelles, et par conséquent, les directives proposées par STROBE sont
43 pertinentes et applicables. Cependant, étant donné que les directives de STROBE ont été conçues
44 pour être appliquées à toutes les études observationnelles, les problèmes spécifiques liés aux
45 données de routine en santé ne sont pas abordés dans STROBE. Un groupe international de
46 scientifiques ayant un intérêt particulier pour l'utilisation des données de santé collectées en
47 routine et des représentants du groupe STROBE se sont rencontrés à Londres en 2012 pour
48 parler de STROBE dans le cadre d'études utilisant des données de routine en santé [14, 15]. Ils
49 ont identifié d'importantes lacunes dans STROBE pour évaluer les études utilisant ce type de
50 données. Ils ont suggéré qu'une extension de STROBE était nécessaire. Ainsi, l'initiative
51 RECORD a été établie sur la base d'un processus collaboratif international ayant abouti à une
52 extension de STROBE pour explorer et traiter les problèmes spécifiques à l'évaluation de la
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

qualité des études réalisées à partir de données collectées en routine. L'initiative RECORD a impliqué plus de 100 parties prenantes au niveau international comprenant des chercheurs, des éditeurs de revues et des utilisateurs de ces données, y compris ceux qui utilisent les résultats des études réalisées à partir de données de routine pour guider leur prise de décision. La méthodologie utilisée pour élaborer les directives RECORD a été détaillée par [16]. Les directrices RECORD sont fondées sur des méthodes reconnues pour élaborer des lignes directrices dans le rapport qualitatif d'études [17]. En bref, les parties prenantes ont été interrogées à deux reprises pour établir et classer par ordre de priorité les thèmes à inclure dans les directives RECORD. Un comité de travail s'est ensuite réuni en présentiel pour rédiger les directives finales. Les intervenants du groupe de travail ont examiné chaque directive et fait des commentaires. La liste finale des items d'évaluation et le présent document d'explication ont été rédigés par les membres du comité de pilotage de RECORD, avec examen et approbation par le groupe de travail. Les membres du comité de pilotage de STROBE ont été impliqués dans la création de RECORD.

Conformément à l'approche STROBE, les directrices RECORD ne sont pas conçues pour recommander des méthodes de recherche, mais pour améliorer la présentation de ces méthodes afin d'assurer les lecteurs, les pairs évaluateurs, les éditeurs de revues et toutes autres personnes d'un moyen d'évaluation de la validité interne et externe des études utilisant des données de santé collectées en routine. En améliorant la qualité des informations rapportées dans les études basées sur des données de santé collectées en routine, nous avons cherché à réduire les ambiguïtés et à respecter les principes du processus scientifique : la découverte, la transparence et la reproductibilité [18].

Éléments de la liste de contrôle RECORD

La liste des items d'évaluation développés par RECORD est fournie dans le Tableau 1. Comme RECORD est fondé sur une extension de STROBE, les énoncés spécifiques à RECORD sont présentés dans une deuxième colonne à celle de STROBE et organisés pour chaque section d'un manuscrit. Nous conseillons aux auteurs d'aborder de manière adéquate chaque item de la liste, mais ne fixons pas d'ordre ou d'emplacement précis dans le manuscrit. Nous avons inclus ci-dessous un texte explicatif pour chaque item de la liste développée par RECORD, selon les sections d'un manuscrit. Dans le cas où aucun élément supplémentaire n'est requis en plus de ceux développés dans STROBE dans les études utilisant des données de santé collectées en routine, les critères d'évaluation sont ceux fournis par l'item de STROBE respectif.

Titre et Résumé

RECORD ITEM 1.1: le type de données utilisé devrait être spécifié dans le titre ou le résumé.

Lorsque cela est possible, le nom des bases de données utilisées devrait être mentionné.

RECORD ITEM 1.2: le cas échéant, les contextes géographique et temporel dans lesquels l'étude a été réalisée devraient être indiqués dans le titre ou le résumé.

RECORD ITEM 1.3: si un chaînage entre différentes bases de données a été réalisé pour l'étude, cela devrait être clairement indiqué dans le titre ou le résumé.

Exemples. Deux exemples d'une bonne documentation de cette section peuvent être illustrés par les articles énumérés ci-dessous :

1. “Perforations and Haemorrhages after Colonoscopy in 2010: A Study Based on Comprehensive French Health Insurance Data (SNIIRAM)” [19].
2. “The Dutch Hospital Standardised Mortality Ratio (HSMR) Method and Cardiac Surgery: Benchmarking in a National Cohort Using Hospital Administration Data versus a Clinical Database” [20].

Explication. Comme il n'existe pas de termes « MeSH » (Medical Subject Heading) permettant l'identification des études réalisées à partir de données de santé collectées en routine, il est important de pouvoir identifier ce type d'études. Compte tenu de la grande diversité des types de données, il ne suffit pas de documenter le fait que des données de routine ont été utilisées. Le type des données de routines devrait être spécifié dans le titre et / ou le résumé. Les exemples de types de données comprennent entre autres les données médico-administratives, les autres données administratives (assurance, registres de naissance / décès et emploi), les registres concernant des maladies, les données de soins primaires, les données issues des dossiers de soins électroniques ainsi que les registres de population. Le fait de nommer les bases de données utilisées est important mais ne remplace pas l'indication du type des sources de données dans le titre ou le résumé.

La région géographique et la période de production des données (contexte temporel) sont des critères d'évaluation déjà présents dans STROBE. Nous suggérons que ces informations soient également documentées dans les sections du titre ou du résumé des manuscrits évalués avec RECORD. La présentation des informations fondamentales concernant la région géographique et la période de production des données devraient évidemment respecter les limites du nombre de mots et tenir compte des questions liées à la confidentialité des données. La région devrait être identifiée au moins par le plus haut niveau géographique pour définir la population étudiée ex., pays, état, province ou région, département ou canton).

De plus, si un chainage entre plusieurs bases de données a été réalisé, il devrait être mentionné dans le titre ou le résumé. Les exemples de termes permettant d'indiquer ce chainage comprennent par exemple « utilisant plusieurs bases de données médico-administratives chaînées » ou « (nom de la base de données) chaînée à (nom de la base de données) ». Les mots « chaîné » ou « chainage » fournissent suffisamment d'informations dans le titre ou le résumé; les détails supplémentaires sur la méthodologie du chainage devraient être fournis dans la section Méthodes des manuscrits évalués.

Introduction

Aucun critère spécifique aux directives de RECORD n'est nécessaire en plus des de ceux établis dans STROBE. Les directives de STROBE recommandent que « des objectifs précis, incluant des hypothèses prédéfinies » soient énoncés dans la section Introduction. L'énoncé des objectifs de recherche spécifiques est essentiel pour la réplique et la traduction de toute recherche observationnelle. Pour les études utilisant des données collectées en routine, les auteurs devraient clarifier si les analyses étaient purement exploratoires dans le but d'identifier de nouvelles relations potentielles dans les données collectées (ex., études génératrices d'hypothèses [21,22]) ou confirmatives dans le but de tester une ou plusieurs

1
2
3 hypothèses [23]. Les auteurs devraient indiquer si leurs hypothèses ont été établies avant ou
4 après l'analyse des données. Ils devraient également indiquer clairement s'il existait un protocole
5 d'étude, comment on peut y accéder le cas échéant, et si l'étude a été enregistrée dans un registre
6 d'étude accessible au public. Étant donné que les points forts et les limites des méthodes utilisées
7 pour effectuer des recherches avec des données recueillies en routine peuvent être controversées,
8 une description claire des objectifs de l'étude est essentielle [23,24]. Il est néanmoins insuffisant
9 de qualifier simplement une étude de descriptive sans préciser si elle vise à établir ou à examiner
10 une hypothèse.
11
12

13 **Méthodes (mise en place)**

14
15
16 Aucun critère RECORD supplémentaire à ceux déjà présent dans les directives de
17 STROBE n'est nécessaire pour évaluer « la description du contexte, des lieux et des dates
18 pertinentes, y compris les périodes de recrutement de la population d'étude, d'exposition, de
19 suivi et de collecte de données ». Les auteurs devraient noter qu'au-delà du type de bases de
20 données dans le titre et / ou le résumé, des informations devraient être fournies afin de permettre
21 aux lecteurs de comprendre le contenu et la validité de la base de données et les raisons
22 originelles pour lesquelles les données ont été collectées. Par exemple, un dossier de soins
23 électronique peut être utilisé par des spécialistes ou des médecins de soins primaires, pour des
24 soins ambulatoires ou hospitaliers, ou par des médecins séniors vs. des étudiants en médecine.
25 Les utilisateurs peuvent être spécialement formés pour une saisie de données exhaustive et
26 reproductible, mais il se peut aussi qu'ils n'aient reçus aucune formation particulière [25]. Les
27 auteurs devraient également décrire comment la population définie par la base de données est
28 comparable avec la population cible, notamment en signalant les critères de sélection potentiels,
29 afin que les lecteurs puissent déterminer si les résultats peuvent être appliqués à la population
30 cible.
31
32
33

34 **Méthodes (participants)**

35
36 RECORD ITEM 6.1: Les méthodes de sélection de la population étudiée (telles que les codes ou
37 les algorithmes utilisés pour identifier les individus) devraient être énumérées en détail. Si cela
38 n'est pas possible, une explication devrait être fournie.
39
40

41 RECORD ITEM 6.2: Toute étude de validation des codes ou des algorithmes utilisés pour
42 sélectionner la population devrait être mise en référence. Si la validation a été effectuée pour
43 cette étude en particulier et n'a pas été publiée par ailleurs, le détail des méthodes utilisées et des
44 résultats de cette validation des données devraient être fournis dans le manuscrit de l'étude
45
46

47 RECORD ITEM 6.3: Si l'étude impliquait le chainage de plusieurs bases de données, il devrait
48 être envisagé l'utilisation d'un diagramme de flux ou d'une autre représentation graphique pour
49 décrire le processus de chainage des données, incluant le nombre d'individus avec les données
50 chaînées à l'issue de chaque étape du chainage.
51

52 **Exemples.** RECORD ITEM 6.1. Un exemple de bonne documentation est fourni dans l'extrait
53 suivant :
54
55
56
57
58
59
60

1
2
3 The OCCC [Ontario Crohn's and Colitis Cohort] uses validated algorithms to identify patients
4 with IBD based on age group. Each of these algorithms was validated in Ontario, in the specific
5 age group to which it was applied, in multiple cohorts, medical practice types, and regions. For
6 children younger than 18 years, the algorithm was defined by whether they underwent
7 colonoscopy or sigmoidoscopy. If they had undergone endoscopy, children required 4 outpatient
8 physician contacts or 2 hospitalizations for IBD within 3 years. If they had not undergone
9 endoscopy, children required 7 outpatient physician contacts or 3 hospitalizations for IBD within
10 3 years. . . . This algorithm correctly identified children with IBD with a sensitivity of. . . [26].

11
12 Cet article fait référence à deux études de validation antérieures d'algorithmes visant à
13 identifier des patients atteints d'une maladie inflammatoire de l'intestin à des âges différents,
14 incluant des mesures de précision du diagnostic. RECORD ITEM 6.2: 1. Dans leur article,
15 Ducharme et ses collègues ont décrit en détail la validation des codes pour identifier les enfants
16 atteints d'intussusception, puis ont utilisé ces codes validés pour décrire l'épidémiologie de la
17 maladie. Les codes impliqués dans l'étude de validation sont décrits dans la figure 2 de l'article
18 [27]. 2. Dans leur article, Benchimol et ses collègues n'ont pas conduit de travail de validation ;
19 cependant, le travail de validation précédemment mené a été mis en référence. La précision des
20 codes diagnostiques utilisés pour développer l'algorithme d'identification a été décrit en détail
21 [26].

22
23 RECORD ITEM 6.3: Les exemples des figures 2, 3 et 4 sur le site Internet de RECORD
24 illustrent plusieurs façons pour décrire le processus de chainage entre des bases de données :

25 Figure 2. Diagramme de Venn pour illustrer un processus de chainage (reproduit avec
26 la permission de Herett et al [28] sur notre site Internet : [http://record-](http://record-statement.org/images/figure2.jpg)
27 [statement.org/images/figure2.jpg](http://record-statement.org/images/figure2.jpg)).

28
29 Figure 3. Diagramme de flux mixte et diagramme de Venn illustrant un processus de
30 chainage (reproduit avec la permission de van Herk-Sukel et al [29] sur notre site Internet :
31 <http://record-statement.org/images/figure3.jpg>).

32
33 Figure 4. Diagramme de chainage combiné avec un diagramme de flux des participants
34 (reproduit avec la permission de Fosbøl et al [30] sur notre site Internet : [http://record-](http://record-statement.org/images/figure4.jpg)
35 [statement.org/images/figure4 .jpg](http://record-statement.org/images/figure4.jpg)).

36
37 **Explication.** RECORD ITEMS 6.1 et 6.2: Documenter la validité des codes / algorithmes
38 d'identification utilisés pour sélectionner la population est nécessaire pour évaluer la
39 transparence dans la présentation d'une recherche observationnelle menée à partir de données de
40 santé collectées en routine. En outre, la documentation des codes / des algorithmes permet à
41 d'autres chercheurs de procéder à une validation externe ou interne.

42
43 Les méthodes utilisées pour identifier les individus d'une étude devraient être
44 explicitement et clairement énoncées, notamment si l'identification est basée sur l'utilisation de
45 codes uniques, d'algorithmes (combinaisons de codes), ou incluant le chainage entre différentes
46 bases de données ou des champs de textes non structurés.

47
48 Le risque de biais de classification dans les études utilisant des données de santé
49 collectées en routine, comme dans beaucoup d'autres études épidémiologiques, peut menacer la
50 validité des résultats de l'étude [31]. Bien que le risque d'erreurs de classification soit amplifié
51 dans des études utilisant des bases de données contenant de larges populations, de telles études
52 offrent l'opportunité d'étudier des maladies rares ou peu communes [32]. La validation des
53 méthodes d'identification a été de plus en plus mise en relief comme étant essentielle pour les
54 études utilisant des données de santé collectées en routine, en particulier pour les codes de
55
56
57
58
59
60

maladies dans les études utilisant des données médico-administratives collectées à des fins de facturation [33]. Les études de validation externes impliquent généralement la comparaison des codes ou des algorithmes utilisés pour identifier les populations étudiées à une norme utilisée comme la référence. Les standards de référence les plus courantes sont les dossiers médicaux, les enquêtes auprès des patients ou des praticiens et les registres cliniques [5,34]. De plus, une validation interne des bases de données peut être entreprise pour comparer les sources de données qui se chevauchent au sein d'une même base de données [35]. Les mesures de précision sont similaires à celles rapportées dans les études de tests diagnostiques, incluant la sensibilité, la spécificité, les valeurs prédictives positives et négatives, ou les coefficients kappa [5,34].

Ainsi, pour les études observationnelles utilisant des données de santé collectées en routine, nous recommandons que les détails de la validation externe ou interne des codes / des algorithmes d'identification soient présentés dans la section Méthodes du manuscrit. Si une ou plusieurs études de validation ont déjà été réalisées, celles-ci doivent être citées dans les références. Si de telles études de validation n'ont pas été réalisées, cela devrait être explicitement indiqué. En outre, une brève discussion sur la précision des méthodes d'identification (en utilisant des termes communs d'exactitude diagnostique) et sur leur fonctionnement dans les sous-populations étudiées devrait être incluse. Si un travail de validation a été réalisé dans le cadre de l'étude observationnelle en question, nous suggérons aux auteurs d'utiliser les directives publiées pour les études de validation [5]. Il est important d'indiquer si la validation s'est produite dans une population cible ou une population définie par une base de données différente de celle sélectionnée pour l'étude en question, car les codes peuvent fonctionner différemment dans différentes populations ou bases de données [36]. En outre, s'il existe des problèmes connus liés à la norme utilisée comme référence à laquelle les données ont été comparées, comme par exemple une incomplétude ou une imprécision, ces problèmes devraient être documentés et discutés comme étant une limite. Les auteurs devraient discuter des implications de l'utilisation de codes / d'algorithmes pour identifier les populations étudiées et les résultats, sur le risque d'erreurs de classification et les impacts potentiels concernant les résultats de l'étude. Il est particulièrement important d'examiner les implications du recours à une étude de validation menée dans une population différente de celle examinée.

RECORD ITEM 6.3: Un diagramme de flux ou une autre présentation graphique peut fournir des informations utiles sur le processus de chaînage et peut simplifier une description potentiellement longue. De telles illustrations peuvent fournir des données clés telles que des informations sur la proportion et les caractéristiques des individus qui ont pu être chaînés et eux qui n'ont pas pu l'être. Les lecteurs devraient être en mesure d'établir la proportion des populations issues des bases de données qui ont été chaînées avec succès ainsi que la représentativité de la population étudiée qui en résulte. Les organigrammes de chaînage peuvent être soit des diagrammes autonomes (ex., diagramme de Venn ou diagrammes de flux), soit combinés avec un diagramme de flux de participants tel que recommandé par les directives STROBE. Comme les graphiques peuvent être fournis dans de nombreux formats, nous ne recommandons pas de modèle spécifique.

Méthodes (variables)

RECORD ITEM 7.1: Une liste complète des codes et des algorithmes utilisés pour classer les expositions, les résultats, les facteurs de confusion et les modificateurs d'effets devrait être

fournie. Si ces codes ou algorithmes ne peuvent pas être documentés, une explication devrait être fournie.

Exemples.

1. Hardelid et ses collègues ont fourni tous les codes dans leur tableau S1 dans le supplément de données 2 [37].
2. Murray et ses collègues ont fourni tous les codes pour les groupes à risque dans leur appendice S1 [38].

Explication. Tout comme pour les codes / les algorithmes utilisés pour identifier la population étudiée, les codes / les algorithmes permettant de classer les expositions, les résultats, les facteurs confondants ou les modificateurs d'effets exposent la recherche à un risque potentiel de biais de classification. Afin de permettre la réplication, l'évaluation et les comparaisons avec d'autres études, nous recommandons qu'une liste de tous les codes diagnostiques, de procédures, de médications ou d'autres codes utilisés pour réaliser l'étude soit fournie dans le manuscrit, dans une annexe en ligne et / ou sur un site Internet externe. Pour les données de routine comprenant des résultats d'enquêtes, les questions de l'enquête devraient être accompagnées de la formulation précise fournie aux sujets de l'étude. Compte tenu du risque de biais de classification dans toutes les recherches, y compris dans les recherches menées à partir de données de santé collectée en routine [31], les auteurs devraient fournir suffisamment de détails pour rendre leur recherche reproductible et le risque de biais visible. Les études de validation peuvent être décrites dans le manuscrit de l'étude ou citées dans les références ou encore accessible en ligne. Comme indiqué ci-dessus, les auteurs devraient indiquer si l'étude de validation a été réalisée dans une population cible ou avec une base de données différente de celle examinée dans l'étude en question.

Nous reconnaissons que dans certaines situations, les chercheurs peuvent être empêchés pour fournir des listes de codes et des algorithmes utilisés dans une publication, car cette information peut être considérée comme exclusive ou protégée par le droit d'auteur, la propriété intellectuelle ou d'autres lois. Par exemple, certains indices d'ajustement sur des comorbidités ont été créés par des entreprises à but lucratif et vendus à des chercheurs dans le cadre de recherches universitaires [39, 40]. Dans ces situations, les auteurs peuvent avoir fait appel à des fournisseurs de données ou à des tiers pour collecter, traiter et / ou chaîner les données. Les auteurs devraient fournir une explication détaillée concernant leur incapacité à fournir des listes de codes ou autres détails sur la façon dont les individus ou les conditions ont été identifiés. Ils devraient également œuvrer pour inclure les coordonnées du groupe détenant des droits de propriété sur ces listes. En outre, les auteurs devraient indiquer comment leur incapacité à fournir cette information peut avoir un impact en termes de réplication et d'évaluation de la recherche. Idéalement, les tiers devraient fournir des informations détaillées sur la manière dont les données ont été collectées, traitées ou chaînées. Une communication améliorée entre les fournisseurs et les utilisateurs de données pourrait être mutuellement bénéfique.

Certains ont trouvé qu'il était nécessaire que les listes de codes représentent la propriété intellectuelle des chercheurs. La publication de listes de codes par d'autres chercheurs que ceux-là même qui les ont développés pourrait priver ces derniers de leur propriété intellectuelle et de la reconnaissance pour avoir créé de telles listes. Nous avons estimé que cette opinion n'était pas compatible avec la norme scientifique définie par la transparence qui permet la reproduction de travaux de recherche. Par conséquent, en dehors de ceux qui sont protégés par la loi ou par contrat, nous recommandons que les listes complètes de codes soient publiées.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Considérant le nombre de mots et les restrictions d'espace dans de nombreuses revues et la longueur potentielle de ces listes de codes / algorithmes, nous reconnaissons que la publication dans un article de journal en format papier peut être impossible. Les informations détaillées peuvent être reportées dans le texte, des tableaux publiés, ainsi que des compléments en ligne sur le site Internet de la revue en annexe, entretenus en permanence par les auteurs ou d'autres personnes, ou déposés dans un référentiel de données tiers (Dryad ou Figshare). Les sections de texte et de référence du manuscrit doivent fournir des informations détaillées sur la façon d'accéder aux listes de codes. Les référentiels de code tels que ClinicalCodes.org sont très prometteurs pour la documentation et la transparence des codes utilisés dans la recherche basée sur les données de santé [41]. Si les listes de codes sont publiées dans des suppléments en ligne sur le site Web de la revue ou sur un site Web externe fourni par les auteurs, le lien devrait être publié dans l'article de revue principal. La publication sur un site Web de journal ou sur PubMed Central (<http://www.ncbi.nlm.nih.gov/pmc/>) augmente la probabilité que le supplément soit disponible tant que le journal est opérationnel. Si la publication sur un site externe privé ou institutionnel est la seule option, nous recommandons que ces listes restent disponibles pendant au moins 10 ans après la publication de l'article de revue. Si l'adresse URL est modifiée, une redirection automatique à partir de l'ancienne adresse Web est requise. Ces mesures permettront aux futurs lecteurs de l'article d'avoir accès aux listes de codes complètes.

En plus des listes de codes fournies dans l'article (ou une annexe en ligne), les auteurs doivent inclure une réflexion sur la possibilité que le choix des codes / algorithmes utilisés dans l'étude conduisent à des biais. Un tel biais pourrait inclure un biais de classification erronée, un biais de constatation et un biais dû à des données manquantes. Si des analyses de sensibilité ont été effectuées en se basant sur de différents ensembles de codes / algorithmes, ceux-ci doivent également être décrits et évalués. Une discussion sur les biais potentiels pourrait également être liée à d'autres parties des listes de contrôle RECORD et STROBE, telles que la sélection des sujets d'étude et la validation des codes (ou leur absence).

Méthodes (méthodes statistiques)

Accès aux données et méthodes de traitement. RECORD ITEM 12.1: Les auteurs doivent décrire à quel degré les chercheurs ont eu accès à la base de données utilisée pour créer la population étudiée.

RECORD ITEM 12.2: Les auteurs doivent fournir des informations sur les méthodes de traitement des données utilisées dans l'étude.

Chainage des données. RECORD ITEM 12.3 : Indiquer si l'étude a inclus un chainage de données au niveau de la personne, au niveau de l'établissement ou d'autres données dans deux bases de données ou plus. Les méthodes de liaison et les méthodes d'évaluation de la qualité du lien devraient être fournies.

Exemples. ITEM 12.1: Les articles suivants décrivent l'accès à un sous-ensemble de la base de données de recherche en médecine générale du Royaume-Uni (GPRD).

1. « The GPRD restricts its data sets to 100,000 individuals for projects funded through the Medical Research Council licence agreement. This restriction mandated a case-control

rather than cohort design to ensure we identified sufficient cases of cancer for each particular symptom... » [42].

2. « A random sample from the General Practice Research Database...was obtained under a Medical Research Council licence for academic institutions » [43].

RECORD ITEM 12.2: Voici un exemple de description de méthodes de traitement de données [44]:

Completeness of common identifiers for linking varied between datasets and by time (identifiers were more complete in recent years). For LabBase2, completeness of identifiers varied by unit (figure 2). For PICANet [Paediatric Intensive Care Audit Network], date of birth and hospital number were 100% complete, and the majority of other identifiers were >98% complete, with the exception of NHS [National Health Service] number (85% complete). For both datasets, cleaning and data preparation were undertaken: NHS or hospital numbers such as “Unknown” or “9999999999” were set to null; generic names (e.g., “Baby,” “Twin 1,” “Infant Of”) were set to null; multiple variables were created for multiple surname and first names; postcodes beginning “ZZ” (indicating no UK postcode) were set to null.

RECORD ITEM 12.3: Les extraits d'articles suivants sont de bons exemples de bons comptes rendus sur le niveau de chaînage de donnée, les techniques et méthodes de chaînage utilisées et les méthodes utilisées pour évaluer la qualité du chaînage :

1. “We linked live birth and fetal death certificates into chronological chains of events that, excluding induced abortions and ectopic pregnancies, constituted the reproductive experience of individual women” [45].
2. Deux articles contiennent d'excellentes descriptions de chaînage qui a été entrepris spécifiquement pour l'étude signalée. [44,45] Dans l'article de Harron et ses collègues [44], une explication détaillée sur le chaînage est fournie avec une démonstration graphique du processus d'appariement. En outre, les méthodes de calcul de la probabilité de chaînage sont décrites : "Match probabilities P(M) (agreement pattern) were calculated to estimate the probability of a match given agreement on a joint set of identifiers. This avoided the assumption of independence between identifiers. Probabilities were derived as the number of links divided by the total number of pairs for each agreement pattern (based on probable links identified in the training datasets). For example, if 378 comparison pairs agreed on date of birth and Soundex but disagreed on sex, and 312 of these were probable links, the match probability for the agreement pattern [1,1,0] was $312/378 = 0.825$ " [44]. L'article d'Adams et ses collègues fournit également une explication détaillée du processus de liaison: "The deterministic linkage consisted of phase I, which entailed six processing steps during which chains were formed and individual (previously unlinked) records were added to chains. Next followed phase n, which entailed multiple passes through the file to combine chains belonging to the same mother" [45].

3. En revanche, si une étude se réfère à des données chaînées antérieurement, la référence faite à un document antérieur serait adéquate comme suit : "Records from both databases were linked to the municipal registries based on date of birth, gender and zip code, and were subsequently linked to each other. The linkage was performed by Statistics Netherlands and is described in previous publications" [20].
4. L'exemple suivant est un exemple de bon rapport sur les caractéristiques des personnes chaînées et non chaînées: "For the purposes of this paper unmatched ISC [Inpatient Statistics Collection] records will be referred to as ISC residuals, unmatched MDC [Midwives Data Collection] records as MDC residuals and linked pairs as matched records... Selected variables that were available on both data sets were compared across three groups-ISC residuals, MDC residuals and matched records" [46].

Explication. RECORD ITEMS 12.1 et 12.2: Des erreurs peuvent survenir si les analystes de données qui ne connaissent pas les nuances de la création de cohorte ou les objectifs de l'étude créent les cohortes d'étude. Par conséquent, la mesure dans laquelle les auteurs ont accès à la base de données devrait être signalée. La description des méthodes de traitement des données à différents stades de l'étude doit inclure celles utilisées pour dépister les données erronées et manquantes, y compris les vérifications des intervalles, les vérifications des doublons et le traitement des mesures répétées [47,48]. D'autres méthodes à signaler pourraient inclure l'évaluation des distributions de fréquences et des tableaux croisés de données et l'exploration graphique ou l'utilisation de méthodes statistiques pour la détection des valeurs aberrantes [49]. Des détails supplémentaires pourraient être fournis sur le diagnostic d'erreur, y compris les définitions de plausibilité et la gestion des erreurs dans l'analyse. Une description claire et transparente des méthodes de traitement des données est importante, car le choix des méthodes pourrait affecter les résultats de l'étude, la répétabilité de l'étude et la reproductibilité des résultats de l'étude [50].

ITEM 12.3: Pour les études de chaînage, nous suggérons de rendre compte du taux estimé de chaînage réussie, de l'utilisation du chaînage déterministe vs probabiliste, de la qualité et du type de variables utilisées pour le chaînage et des résultats de toute validation de chaînage. Si le chaînage des entrées à travers les bases de données a été réalisé spécifiquement pour l'étude, les méthodes de chaînage et d'évaluation de la qualité du chaînage devraient être rapportées, y compris les informations à propos de la personne qui a effectué le chaînage. Si possible, des détails doivent être fournis sur les variables de blocage, l'exhaustivité des variables de chaînage, les règles de chaînage, les seuils et la révision manuelle [44]. Si le chaînage a été réalisé avant l'étude (pour des études antérieures ou pour un usage général) ou si le chaînage de données a été effectué par un fournisseur externe, tel qu'un centre de chaînage de données, une référence est nécessaire décrivant la ressource de données et les méthodes du chaînage.

Les données décrivant les méthodes de chaînage et évaluant leur succès sont essentielles pour permettre au lecteur d'évaluer l'impact de toute erreur de chaînage et des biais associés [51]. Plus précisément, le lecteur devrait savoir si le type de chaînage utilisé était déterministe et / ou probabiliste, afin de déterminer si le chaînage pourrait être affecté par de fausses appariements ou d'appariement manquées. Le chaînage déterministe est utile lorsqu'un identifiant unique est disponible dans les différentes sources de données. Lorsqu'un tel identifiant n'est pas disponible, une description des règles de chaînage d'enregistrements appliquées (ou des clés de chaînage statistiques) est essentielle. En revanche, le chaînage probabiliste utilise

1
2
3 plusieurs identifiants, parfois avec des poids différents, et les correspondances sont considérées
4 comme présentes au-dessus d'un seuil spécifique. Des méthodes mixtes peuvent également être
5 utilisées. Par exemple, un chaînage déterministe peut être utilisé pour certains entrées, et un
6 chaînage probabiliste peut être appliqué lorsque des identifiants uniques ne sont pas disponibles
7 pour d'autres entrées. Le biais de chaînage se produit lorsque des associations sont présentes
8 entre la probabilité d'erreur de liaison (p. ex., les correspondances fausses et manquantes) et les
9 variables d'intérêt. Par exemple, les taux de liaison peuvent varier selon les caractéristiques du
10 patient, p. ex., l'âge, le sexe et l'état de santé. Même de petites erreurs dans le processus de
11 chaînage peuvent causer un biais et conduire à des résultats qui peuvent surestimer ou sous-
12 estimer les associations étudiées [52]. Les auteurs devraient signaler une erreur de chaînage en
13 utilisant des approches standard, y compris des comparaisons avec des normes de référence ou
14 des ensembles de données de référence, des analyses de sensibilité et en comparant les
15 caractéristiques des données chaînées et non chaînées [53]. La signalisation d'une erreur de
16 liaison permet au lecteur de déterminer la qualité du chaînage et la possibilité de biais lié aux
17 erreurs de liaison.
18
19
20

21 **Résultats (Participants)**

22 RECORD ITEM 13.1: Décrire en détail la sélection des personnes incluses dans l'étude (c'est-à-
23 dire la sélection de la population étudiée), y compris le filtrage basé sur la qualité des données, la
24 disponibilité des données et le chaînage. La sélection des personnes incluses peut être décrite
25 dans le texte et / ou au moyen d'un diagramme de flux
26
27

28 **Exemple.** Un exemple de bon compte-rendu est donné dans l'extrait suivant :

29
30
31 We identified 161,401 Medicare beneficiaries given a diagnosis of one or more cases of
32 can- cer of the lung and bronchus in the SEER [Surveillance, Epidemiology, and End
33 Results] registries between 1998 and 2007. Among these patients, we identified a total of
34 163,379 separate diagnoses of incident lung cancer. (Some patients had two cases of
35 primary lung cancer separated by more than a year during the study period). Fig 1 shows
36 the derivation of the final cohort of 46,544 patients with 46,935 cases of NSCLC [non-
37 small cell lung cancer] [54]. (Voir la figure 5 pour l'exemple de diagramme de flux,
38 disponible sur <http://record-statement.org/images/figure5.jpg>.)
39
40

41 **Explication.** Les auteurs doivent fournir une description claire de la population issue de la base
42 de données originale des données de santé collectées de manière routinière, car les différences
43 entre la population étudiée et la population de la base de données doivent être documentées pour
44 permettre l'application des résultats (voir aussi RECORD item 6.1). Les chercheurs utilisant des
45 sources de données collectées de manière routinière limitent souvent leur population étudiée en
46 fonction de facteurs tels que la qualité des données disponibles. Par exemple, ils peuvent
47 restreindre la période d'étude à un temps pendant lequel la qualité des données est jugée
48 acceptable, ce qui entraîne l'exclusion de participants potentiels. Des études peuvent exclure des
49 pratiques médicales avec des entrées de dossiers de santé électroniques incohérentes ou attendre
50 que ces pratiques deviennent cohérentes [38,55]. La population étudiée peut également être
51 restreinte en fonction de la disponibilité de données. Par exemple, dans les études utilisant les
52 données de l'assurance-maladie des États-Unis, les bénéficiaires actuellement inscrits dans un
53 organisme de maintien de la santé sont souvent exclus en raison de l'absence d'entrées
54
55
56
57
58
59
60

1
2
3 d'événements cliniques [54,56]. Lors de l'utilisation de sources de données dans lesquelles
4 l'éligibilité fluctue au fil du temps (p. ex., les bases de données d'assurance), les chercheurs
5 doivent préciser clairement comment l'éligibilité a été définie et comment les modifications
6 d'éligibilité ont été gérées dans leur étude. Si une étude utilise des données de routine chaînées,
7 la population de l'étude est fréquemment réduite en se limitant aux individus pour lesquels des
8 données chaînées sont disponibles [57]. Des cohortes fortement restreintes peuvent également
9 être utilisées pour des raisons méthodologiques afin d'éliminer certains facteurs confondants.

10
11 Ainsi, les étapes suivies pour obtenir les populations étudiées définitives, les critères
12 d'inclusion et d'exclusion, et l'inclusion et l'exclusion des participants aux différentes étapes de la
13 création et de l'analyse des cohortes doivent être clairement définies soit dans le texte, soit sous
14 forme d'une représentation schématique. Les populations étudiées peuvent être dégagées à l'aide
15 de codes et / ou d'algorithmes différents (voir l'item 6.1 du RECORD), ce qui peut avoir un
16 impact sur la population étudiée au fil du temps [58,59]. Certaines études peuvent également
17 avoir utilisé plusieurs définitions de cas plus ou moins sensibles / spécifiques, ce qui peut avoir
18 un impact sur les analyses ultérieures. La délimitation de ces étapes est importante pour évaluer
19 la validité externe des résultats de l'étude et, dans certaines circonstances, pour évaluer un
20 éventuel biais de sélection. Des analyses de sensibilité peuvent être présentées pour évaluer
21 l'impact potentiel de l'absence de données et de la représentativité de la population étudiée. Le
22 fait de fournir des informations sur la sélection de la/des population (s) étudiée (s) à partir de la
23 base de données initiale permet également de répliquer l'étude. Des analyses subsidiaires peuvent
24 avoir été effectuées sur différentes populations d'étude et peuvent potentiellement être présentées
25 dans des annexes en ligne.

26 27 28 29 30 **Discussion (limites)**

31
32 RECORD ITEM 19.1: Discuter les implications de l'utilisation de données qui n'ont pas été
33 créées ou collectées pour répondre aux questions de recherche spécifiques. Inclure une
34 discussion sur les biais de classification erronée, les facteurs confondants non mesurés, les
35 données manquantes et l'évolution de l'admissibilité au fil du temps, en ce qui a trait à l'étude
36 faisant l'objet du rapport.

37
38
39 **Exemples.** Les articles suivants décrivent les limites associées à l'utilisation des données
40 administratives :

- 41 1. "Third, this study was a retrospective, claims-based analysis. Only PET [positron
42 emission tomography] scans paid for by Medicare could be detected in the analysis. To
43 minimize the proportion of missed claims, all analyses were limited to Medicare
44 beneficiaries with both Medicare Part A and Part B coverage and no enrollment in
45 managed care or Medicare Part C for the 12 months before and after diagnosis. Fourth,
46 patients in the SEER registry are more likely to be nonwhite, to live in areas with less
47 poverty, and to live in urban areas, which may limit the generalizability of the findings.
48 Fifth, during the study period, disease stage was based on SEER data obtained over 4
49 months or until first surgery. In 2004, data collection for SEER changed to the
50 collaborative staging system. It is unclear how our results would differ with this newer
51 approach" [54].
52
53
54
55
56
57
58
59
60

- 1
2
3 2. “Despite several strengths of the SEER-Medicare data, including a comparatively large
4 sample size, generalizability to the US population, and detailed information on
5 prescriptions, our study was limited by the lack of laboratory data on cholesterol,
6 triglyceride, and glucose levels that would have informed the extent of metabolic
7 disturbances in the population. . .thus having laboratory-based data could have reduced
8 residual confounding by severity of metabolic disease. We also lacked more granular data
9 on cancer progression, which could have confounded the association between statin use
10 and death, given that statin treatment may be withheld or discontinued in patients with
11 short expected survival time” [60].
12
13

14
15 **Explication.** Les données de santé collectées de manière routinière ne sont généralement pas
16 collectées en se basant sur une question de recherche spécifique développée a priori, et les
17 raisons qui motivent la collecte de données peuvent varier. De nombreux domaines potentiels de
18 biais, y compris toutes les sources habituelles de biais associées à la recherche observationnelle
19 mais aussi plus spécifiques à la recherche observationnelle utilisant des données de routine,
20 menacent les conclusions des chercheurs. Les auteurs devraient considérer les éléments suivants
21 comme des sources potentielles de biais : (1) codes ou algorithmes pour identifier les populations
22 étudiées, les critères de jugement, les facteurs confondants ou les modificateurs d'effets (biais de
23 classification erronée); (2) variables manquantes (facteurs confondants non mesurés); (3)
24 données manquantes; et (4) les changements d'admissibilité au fil du temps.
25

26 La logique sous-jacente à la collecte de données de routine peut affecter la qualité et
27 l'applicabilité des données aux questions de recherche examinées. Par exemple, les registres
28 utilisés pour les analyses rétrospectives peuvent mettre en œuvre un meilleur contrôle de la
29 qualité que les organisations collectant d'autres types de données collectées de manière
30 routinière, bien que cela puisse varier. De même, certaines données administratives sont
31 soumises à un contrôle de qualité minutieux, alors que d'autres ne le sont pas. Les données
32 administratives sont particulièrement sujettes au codage abusif ou opportuniste. Par exemple,
33 lorsque le remboursement hospitalier est basé sur la complexité de la gamme de cas, les hôpitaux
34 pourraient maximiser le remboursement en appliquant généreusement des codes de maladie plus
35 complexes aux dossiers des patients [61]. De plus, des changements dans les stratégies de codage
36 peuvent avoir une incidence sur la validité ou la consistance des données. Par exemple,
37 l'introduction de codes d'incitation à la facturation du fournisseur peut modifier la probabilité
38 d'utilisation d'un code au fil du temps [62,63]. D'autres codes peuvent être évités en raison de la
39 stigmatisation du patient ou des pénalités du fournisseur [64]. De plus, des changements dans les
40 versions des systèmes de classification des codes (p. ex., de la Classification internationale des
41 maladies (CIM) -9 à la CIM -10) peuvent altérer la validité de la constatation utilisant des
42 données codées [65,66]. La variation de la pratique clinique dans les hôpitaux et selon les
43 populations peut entraîner une localisation des examens de laboratoire dans des endroits et / ou
44 des pratiques spécifiques, ce qui peut avoir un impact sur un algorithme diagnostique. Si l'une de
45 ces sources potentielles de biais de classification se présentent, elles devraient être considérées
46 comme des limites à l'étude.
47

48 Les facteurs confondants non mesurés sont définies comme étant des facteurs associés à des
49 variables non incluses dans les données de l'étude, ce qui entraîne un biais de confusion résiduel
50 [67]. Bien qu'ils représentent une source potentielle de biais dans toutes les recherches par
51 observation, ils sont particulièrement prédominants dans les études utilisant des données
52 collectées de façon routinière.
53
54
55
56
57
58
59
60

1
2
3 L'analyse pourrait nécessiter des variables qui n'ont pas été prises en compte lors de la
4 planification des bases de données ou lors de la collecte des données. Diverses méthodes ont été
5 proposées pour traiter cette source potentielle de biais [68-71], y compris les scores de
6 propension. Cependant, les analyses de score de propension, comme les analyses de régression
7 standard et l'appariement, ne peuvent garantir qu'un équilibre des participants à l'étude sur les
8 variables disponibles dans les données. Un type particulier de confusion non mesurée est la
9 confusion par indication ; c'est souvent un problème lorsqu'on examine l'efficacité et l'innocuité
10 des traitements (médicamenteux) en utilisant des données de routine. Par conséquent, le
11 pronostic des personnes recevant le traitement (médicamenteux) peut être mieux ou pire que
12 celui des autres, mais les données sur le pronostic et / ou la sévérité de la maladie sous-jacente
13 peuvent ne pas être disponibles dans les données [72]. Ces questions devraient être discutées par
14 les auteurs, et les méthodes utilisées pour en tenir compte (si possible) devraient être signalées.

15
16 Les données manquantes sont problématiques pour toute recherche observationnelle et
17 ont été traitées dans l'encadré 6 de l'article explicatif STROBE [10]. Les données manquantes
18 sont un problème particulier pour les données collectées de façon routinière, car les chercheurs
19 ne peuvent pas contrôler la collecte de données [73]. Les données manquantes peuvent entraîner
20 un biais de sélection s'il existe des valeurs manquantes dans les variables utilisées pour définir la
21 cohorte de l'étude ou des identifiants manquants qui entravent le chaînage des entrées, surtout si
22 les données manquantes existent de manière non aléatoire. Les variables manquantes créent des
23 défis similaires. Les auteurs devraient déterminer les variables manquantes suspectées de causer
24 un biais de confusion non mesuré, la raison pour laquelle ces variables manquaient, comment
25 cela a pu affecter les résultats des études et les méthodes utilisées pour ajuster les variables
26 manquantes. Par exemple, le tabagisme a un fort effet sur la sévérité de la maladie de Crohn et a
27 été associé à l'évolution de cette maladie. Cependant, le statut tabagique est rarement inclus dans
28 les données administratives sur la santé. Dans une étude évaluant l'association entre le statut
29 socio-économique et les résultats de la maladie de Crohn, le statut tabagique a été abordé comme
30 un facteur de confusion potentiel non mesuré [74]. Fréquemment, les données manquantes /
31 variables manquantes ne sont découvertes qu'après le début de la recherche utilisant des données
32 de santé collectées de manière routinière, ce qui oblige les chercheurs à s'écarter de leur
33 protocole de recherche original. Les détails de la déviation par rapport au protocole, quelle que
34 soit sa raison, doivent toujours être indiqués. Les raisons de la déviation et les implications sur la
35 recherche et les conclusions devraient être abordées.

36
37 Une autre limitation potentielle importante est l'évolution des pratiques de codage ou
38 des critères d'admissibilité résultant d'une modification de la composition de la base de données,
39 de la population étudiée ou des deux au fil du temps. La définition de la population de base de
40 données peut changer dans un certain nombre de circonstances, comme par exemple si les
41 cabinets de recrutement cessent de collaborer avec la base de données, changent de logiciel ou
42 modifient les critères d'inscription dans la base de données, par exemple, registre. La population
43 étudiée dans les sources de données administratives (p. ex., les bases de données d'assurance)
44 peut changer si l'éligibilité des personnes n'est pas constante dans le temps, en raison de
45 changements d'emploi, de statut de résidence ou de fournisseur de soins médicaux. Un
46 changement dans la façon dont les entrées sont codées (p. ex., codage abusif ou changements
47 dans les systèmes de codage, comme décrit ci-dessus) peut modifier la population de l'étude [63,
48 75, 76]. Lorsqu'ils abordent les limites, les auteurs doivent expliquer comment le changement
49 d'éligibilité a été traité dans l'analyse afin que le lecteur puisse évaluer la possibilité d'un biais.

1
2
3 Comme indiqué par les directives STROBE, la discussion devrait inclure l'orientation et
4 l'ampleur de tout biais potentiel et les efforts déployés pour remédier à ce biais.
5

6 **Autre Informations**

7
8
9 RECORD ITEM 22.1: Les auteurs doivent fournir des informations sur la façon d'accéder à toute
10 information supplémentaire telle que le protocole d'étude, les données brutes ou le code de
11 programmation.
12

13 **Exemples.**

- 14
15
16 1. L'article de Taljaard et ses collègues représentent le protocole de recherche complet pour
17 un étude basée sur CCHS (the Canadian Community Health Survey) [77].
18
- 19 2. Dans son article, Guttman et ses collègues invitent des demandes pour le protocole
20 d'étude : "Data sharing : The technical appendix, dataset creation plan/protocol, and
21 statistical code are available from the corresponding author at [email address]" [78].
22
23

24 **Explication.** Nous soutenons fortement la diffusion d'informations détaillées sur les méthodes de
25 l'étude et les résultats. Lorsque cela est possible, nous encourageons la publication préalable ou
26 simultanée du protocole d'étude, des résultats de données brutes et, le cas échéant, du code de
27 programmation. Cette information est utile aux évaluateurs et aux lecteurs pour évaluer la
28 validité des résultats de l'étude. Un certain nombre d'opportunités sont à la disposition des
29 chercheurs pour la publication ouverte de ces données. Ceux-ci comprennent des documents
30 supplémentaires en ligne, des sites Web personnels, des sites Web institutionnels, des sites de
31 médias sociaux à vocation scientifique (ResearchGate.net et Academia.edu), des répertoires de
32 données (par exemple, Dryad ou Figshare) ou des sites Web gouvernementaux ouverts. Nous
33 reconnaissons que certaines organisations de recherche, sociétés, institutions ou lois peuvent
34 interdire ou restreindre la disponibilité libre de ces informations. Bien qu'une discussion sur la
35 propriété et l'utilisation de cette propriété intellectuelle soit hors du champ d'application des
36 directives RECORD, la publication de ces données doit toujours être effectuée dans le respect
37 des directives légales et éthiques de l'environnement institutionnel des chercheurs, avec l'aide des
38 éditeurs. Cette information serait également utile à d'autres chercheurs qui souhaiteraient accéder
39 à ces données pour reproduire ou développer la recherche décrite dans le manuscrit. Quel que
40 soit le format ou l'étendue des informations supplémentaires disponibles, nous recommandons
41 que la référence à l'emplacement de ces informations soit clairement indiquée dans le manuscrit.
42
43
44
45

46 **Discussion**

47 Les directives RECORD sont spécifiques à la recherche observationnelle réalisée à partir de
48 données de santé collectées de manière routinière et servent à compléter et non à remplacer les
49 directives STROBE. RECORD a été créé pour guider les auteurs, les éditeurs de revues, les pairs
50 évaluateurs et les autres parties prenantes afin d'encourager la transparence et l'exhaustivité des
51 rapports de recherche menés en utilisant des données de santé collectées de manière routinière.
52 La liste de contrôle est destinée à être utilisée par tout chercheur utilisant de telles données, et
53 nous encourageons une large diffusion auprès de toutes les parties intéressées. Nous prévoyons
54
55
56
57
58
59
60

1
2
3 que l'approbation et la mise en œuvre de RECORD par les revues amélioreront la transparence
4 des rapports de recherche utilisant des données de santé collectées de manière routinière.
5
6

7 **Limites**

8 L'application de STROBE et RECORD est uniquement destinée aux études de recherche
9 observationnelles. Cependant, des données de santé collectées de manière routinière sont parfois
10 utilisées pour des recherches menées avec d'autres études, telles que essais randomisés en cluster
11 pour l'évaluation des systèmes de santé. De plus, le chaînage de données provenant d'essais
12 randomisés à des données administratives peut être utilisé pour le suivi à long terme des
13 résultats, et les études associées ne seraient pas considérées comme des observations. À mesure
14 que le domaine évolue, nous prévoyons étendre RECORD à d'autres plan de recherche en
15 utilisant des méthodes rigoureusement similaires.
16

17 Bien que RECORD représente notre meilleure tentative pour refléter l'intérêt et les
18 priorités des parties prenantes, nous reconnaissons que les méthodes utilisées pour mener des
19 recherches utilisant des données de santé collectées régulièrement évoluent rapidement et que la
20 disponibilité des types de données pour ces recherches augmente. Par exemple, les applications
21 de santé mobiles (mHealth apps) deviennent de plus en plus disponibles pour les smartphones et
22 les technologies portables. Bien que la recherche soit actuellement limitée à l'aide de ces sources
23 de données, nous prévoyons une croissance rapide de l'utilisation de ces données dans un proche
24 avenir, et de nouvelles méthodologies seront créées pour gérer cette ressource. En outre, le
25 comité de travail a décidé de se concentrer sur les données de santé et non sur toutes les sources
26 de données utilisées pour mener des recherches sur la santé (p. ex., données environnementales,
27 données financières, etc.). Par conséquent, la liste de contrôle de RECORD pourrait ne pas
28 refléter les thèmes qui deviendront importants à l'avenir, et une révision pourrait être nécessaire à
29 un moment donné.
30
31

32 Des efforts importants ont été déployés pour inclure une large représentation des parties
33 prenantes dans la création de ces lignes directrices. Nous avons recruté des parties prenantes à
34 travers des appels ouverts et des invitations ciblées en utilisant une variété de canaux [16].
35 Cependant, la représentation des parties prenantes provenait principalement de régions menant
36 des recherches utilisant des données de santé collectées de manière routinière, avec seulement
37 quelques représentants de pays en développement et de pays non anglophones. Néanmoins, nous
38 croyons que le groupe des intervenants était représentatif de la communauté actuelle des
39 chercheurs et des utilisateurs de connaissances générées. Bien que de nombreuses contributions
40 aient été obtenues grâce aux enquêtes et retours du groupe de parties prenantes, la faisabilité
41 dictait que les déclarations soient rédigées par un comité de travail plus restreint composé de 19
42 membres qui se rencontraient en personne [17]. À l'avenir, la technologie et les médias sociaux
43 pourraient permettre une participation plus active de groupes plus importants aux réunions des
44 comités de travail.
45
46
47

48 **Orientations futures et engagement communautaire**

49 Au fur et à mesure que la disponibilité des données de santé collectées de manière routinière
50 augmente, nous nous attendons à une plus grande implication des chercheurs des régions dans
51 lesquelles ces données ne sont pas actuellement accessibles. Par le biais du site Web record-
52 statement.org et du babillard électronique, nous nous attendons à des commentaires continus et à
53 des discussions sur le document RECORD de la part des parties intéressées, ce qui pourrait
54
55
56
57
58
59
60

entraîner des révisions officielles à l'avenir. Grâce à cette communauté en ligne, RECORD deviendra un document vivant qui peut s'adapter aux changements sur le terrain.

La publication d'une ligne directrice et l'approbation des revues ne sont pas suffisantes pour améliorer les rapports de recherche [80]. La manière de mettre en œuvre les lignes directrices par les chercheurs, les revues et les pairs évaluateurs est d'une importance cruciale pour que RECORD ait un impact mesurable [81]. Par conséquent, le forum en ligne inclura un forum de discussion sur la mise en œuvre. Nous encourageons également l'évaluation de l'impact de RECORD sur les rapports sur le terrain afin de garantir que les directives fournissent des avantages mesurables.

Conclusions

La déclaration RECORD applique les critères STROBE aux études observationnelles menées à l'aide de données de santé collectées de manière routinière. Grâce à la contribution de la communauté de la recherche et de l'édition, nous avons créé des lignes directrices pour la présentation de rapports sous la forme d'une liste de contrôle et du présent document explicatif. Il a été démontré que les lignes directrices pour la présentation des rapports améliorent le compte rendu de la recherche, ce qui permet aux consommateurs de la recherche d'être conscients des points forts, des limites et de l'exactitude des conclusions [12, 82-84]. Bien que nous prévoyions que RECORD évoluera avec l'évolution des méthodes de recherche sur le terrain, ces lignes directrices aideront à faciliter l'établissement de rapports adéquats sur la recherche au cours des prochaines années. Grâce à la mise en œuvre par les auteurs, les éditeurs de revues et les pairs évaluateurs, nous prévoyons que RECORD se traduira par la transparence, la reproductibilité et l'exhaustivité des rapports de la recherche menée en utilisant des données de santé collectées de manière routinière.

Renseignements à l'appui

Table S1. Liste des parties prenantes ayant participé aux enquêtes. (XLSX)

Remerciements

Ce travail a été approuvé par le Comité d'éthique de la recherche du Centre hospitalier pour enfants de l'est de l'Ontario. Les auteurs remercient les parties prenantes qui ont participé aux enquêtes d'avoir accordé la priorité aux thèmes à inclure dans la liste de contrôle (tableau S1). Nous sommes également reconnaissants aux membres du groupe d'initiative STROBE, qui ont guidé et soutenu la création de RECORD. Les auteurs remercient chaleureusement Pauline Quach et Danielle Birman, coordonnatrices de la recherche à RECORD, ainsi que Andrew Perlmutar, concepteur du site Web et administrateur de record-statement.org. Les auteurs sont également reconnaissants pour la contribution de toutes les parties prenantes qui ont participé à l'élaboration de ces lignes directrices.

Membres du comité de travail RECORD: Douglas Altman (Centre for Statistics in Medicine, Oxford University), Nicholas de Klerk (University of Western Australia), Lars G. Hemkens (University Hospital Basel), David Henry (University of Toronto and Institute for Clinical Evaluative Sciences, Toronto), Jean-Marie Januel (University of Lausanne), Marie-Annick Le Pogam (Institute of Social and Preventive Medicine, University Hospital of Lausanne), Douglas Manuel (Ottawa Hospital Research Institute, University of Ottawa), Kirsten Patrick (rédactrice,

1
2
3 Canadian Medical Association Journal [CMAJ]), Pablo Perel (London School of Hygiene and
4 Tropical Medicine), Patrick S. Romano (University of California, Davis, co-rédacteur en chef,
5 Recherche sur les services de santé), Peter Tugwell (University of Ottawa, rédacteur en chef,
6 Journal of Clinical Epidemiology), Joan Warren (National Institutes of Health / National Cancer
7 Institute), Wim Weber (rédacteur en chef européen, BMJ) et Margaret Winker (anciennement
8 éditeur, PLOS Medicine; secrétaire actuelle, World Association of Medical Editors).
9
10

11 **Contributions des auteurs**

12 Rédaction de la première version du manuscrit : EIB SML. Contribution à la rédaction du
13 manuscrit : EIB LS AG KH DM IP HTS EvE SML Membres du comité de travail RECORD.
14 Approbation des résultats et conclusions du manuscrit : BEI LS AG KH DM IP HTS EvE SML
15 Membres du comité de travail RECORD
16
17

18 **Encadré 1.** Définitions des termes de la population (sources, bases de données et populations
19 étudiées).

20 Il existe trois niveaux dans la hiérarchie des populations, qui sont pertinents dans les études
21 utilisant des données collectées de manière routinière et qui seront mentionnés tout au long du
22 document. Ces populations comprennent la population source, qui représente celle à partir de
23 laquelle la population de la base de données est dérivée et donc de laquelle les chercheurs
24 relèvent des déductions; la population de base de données, qui est dérivée de la population source
25 et comprend des personnes avec des données incluses dans la source de données; et la population
26 étudiée, identifiée à partir de la population de la base de données par les chercheurs utilisant des
27 codes et des algorithmes (Fig 1) [7]. Par exemple, dans le cas du « Clinical Practice Research
28 Datalink » (CPRD), la population source comprend toutes les personnes qui fréquentent des
29 médecins généralistes au Royaume-Uni. La population de base de données comprend les
30 individus inclus dans le CPRD, alors que la population étudiée comprend ceux choisis du CPRD
31 en utilisant des codes et des algorithmes qui seront décrits dans l'étude spécifique.
32
33
34

35 **Figure 1.** Hiérarchie des populations dans les études menées en utilisant des sources de données
36 collectées de manière routinière.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Confidential

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Liste de contrôle RECORD - étendue de la déclaration STROBE - des éléments qui devraient être inclus pour rendre compte des études observationnelles utilisant des données de santé collectées de manière routinière.

	Numéro d'item	STROBE items	Emplacement dans le manuscrit	RECORD items	Emplacement dans le manuscrit
Titre et résumé					
	1	(a) Indiquer dans le titre ou le résumé le type d'étude avec un terme couramment utilisé (b) Fournir un résumé informatif et équilibré de ce qui a été fait et ce qui a été trouvé		RECORD ITEM 1.1: Le type de données utilisé devrait être spécifié dans le titre ou le résumé. Lorsque cela est possible, le nom des bases de données utilisées devrait être mentionné. RECORD ITEM 1.2: Le cas échéant, les contextes géographique et temporel dans lesquels l'étude a été réalisé devraient être indiqués dans le titre ou le résumé. RECORD ITEM 1.3: Si un chaînage entre différentes bases de données a été réalisé pour l'étude, cela devrait être clairement indiqué dans le titre ou le résumé.	
Introduction					
Contexte/ Justification	2	Expliquer le contexte scientifique et la logique sous-jacente dans l'étude rapportée			
Objectifs	3	Énoncer les objectifs spécifiques, y compris toutes les hypothèses préétablies			
Méthodes					
Conception de l'étude	4	Présenter les éléments clés du type d'étude au début du document			

Contexte	5	Décrire le contexte, les lieux et les dates pertinentes, y compris les périodes de recrutement, d'exposition, de suivi et de collecte de données			
Participants	6	<p>(a) <i>Étude de cohorte</i> - Donner les critères d'admissibilité ainsi que les sources et les méthodes de sélection des participants. Décrire les méthodes de suivi</p> <p><i>Étude cas-témoins</i> - Donner les critères d'admissibilité, ainsi que les sources et les méthodes de détermination des cas et de contrôle. Donner la justification du choix des cas et des contrôles</p> <p><i>Étude transversale</i> - Donner les critères d'admissibilité, ainsi que les sources et les méthodes de sélection des participants</p> <p>(b) <i>Étude de cohorte</i> - Pour les études appariées, indiquer les critères d'appariement et le nombre de sujets exposés et non exposés</p> <p><i>Étude cas-témoins</i> - Pour les études appariées, donner des critères de correspondance et</p>		<p>RECORD ITEM 6.1: Les méthodes de sélection de la population étudiée (telles que les codes ou les algorithmes utilisés pour identifier les individus) devraient être énumérées en détail. Si cela n'est pas possible, une explication devrait être fournie.</p> <p>RECORD ITEM 6.2: Toute étude de validation des codes ou des algorithmes utilisés pour sélectionner la population devrait être mise en référence. Si la validation a été effectuée pour cette étude en particulier et n'a pas été publiée par ailleurs, le détail des méthodes utilisées et des résultats de cette validation des données devraient être fournis dans le manuscrit de l'étude.</p> <p>RECORD ITEM 6.3: Si l'étude impliquait le chaînage de plusieurs bases de données, il devrait être envisagé l'utilisation d'un diagramme de flux ou d'une autre représentation graphique pour décrire le processus de chaînage des données, incluant le nombre d'individus avec les données chaînées à l'issue de chaque étape du chaînage.</p> <p>.</p>	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

		le nombre de contrôles par cas			
Variables	7	Définir clairement tous les résultats, les expositions, les prédicteurs, les facteurs confondants potentiels et les facteurs d'influence. Donner des critères de diagnostic, le cas échéant.		RECORD 7.1: Une liste complète des codes et des algorithmes utilisés pour classer les expositions, les résultats, les prédicteurs, les facteurs confondants potentiels et les facteurs d'influence devrait être fournie. Si ceux-ci ne peuvent pas être signalés, une explication devrait être fournie.	
Sources de données / mesures	8	Pour chaque variable d'intérêt, donnez les sources de données et les détails des méthodes d'évaluation (mesure). Décrire la comparabilité des méthodes d'évaluation s'il y a plus d'un groupe			
Biais	9	Décrivez les efforts déployés pour traiter les sources potentielles de biais			
Taille de l'étude	10	Décrire comment la taille de l'échantillon dans l'étude a été déterminée			
Variabes quantitatives	11	Expliquer comment les variables quantitatives ont été traitées dans les analyses. Décrire, s'il y a lieu, quels groupements ont été choisis et pourquoi			
Méthodes statistiques	12	(a) Décrire toutes les méthodes statistiques, y compris celles utilisées pour			

		<p>contrôler les facteurs confondants</p> <p>(b) Décrire toutes les méthodes utilisées pour examiner les sous- groupes et les interactions</p> <p>(c) Expliquer comment les données manquantes ont été traitées</p> <p>(d) <i>Étude de cohorte</i> – Le cas échéant, expliquer comment la perte de suivi a été abordée <i>Étude cas-témoin</i> – Le cas échéant, expliquer comment l'appariement des cas et des témoins a été réalisé <i>Étude transversale</i> –Le cas échéant, décrire les méthodes d'analyse tenant compte de la stratégie d'échantillonnage</p> <p>(e) Décrire toutes les analyses de sensibilité</p>			
Accès aux données et méthodes de traitement		..		<p>RECORD 12.1: Les auteurs doivent décrire à quel degré les chercheurs ont eu accès à la base de données utilisée pour créer la population étudiée.</p> <p>RECORD 12.2: Les auteurs doivent fournir des informations sur les méthodes de traitement des données utilisées dans l'étude.</p>	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Chainage des données		..		RECORD 12.3: Indiquer si l'étude a inclus un chainage de données au niveau de la personne, au niveau de l'établissement ou d'autres données dans deux bases de données ou plus. Les méthodes de liaison et les méthodes d'évaluation de la qualité du lien devraient être fournies.	
Résultats					
Participants	13	<p>(a) Indiquer le nombre d'individus à chaque étape de l'étude (par exemple, les nombres potentiellement admissibles, examinés pour l'admissibilité, confirmés admissibles, inclus dans l'étude, complétant le suivi et analysés)</p> <p>(b) indiquer les raisons de la non-participation à chaque étape.</p> <p>(c) Envisager l'utilisation d'un diagramme de flux</p>		RECORD 13.1: Décrire en détail la sélection des personnes incluses dans l'étude (c'est-à-dire la sélection de la population étudiée), y compris le filtrage basé sur la qualité des données, la disponibilité des données et le chainage. La sélection des personnes incluses peut être décrite dans le texte et / ou au moyen d'un diagramme de flux	
Données descriptives	14	<p>(a) Indiquer les caractéristiques des participants à l'étude (par exemple, démographiques, cliniques, sociales) et les informations sur les expositions et les facteurs confondants potentiels</p> <p>(b) Indiquer le nombre de</p>			

		<p>participants avec des données manquantes pour chaque variable d'intérêt</p> <p>(c) <i>Étude de cohorte</i>- Résumer le temps de suivi (par exemple, la moyenne et le temps total)</p>			
Données de résultats	15	<p><i>Étude de cohorte</i> - Rapporter le nombre d'évènements ou les indicateurs mesurés au fil du temps</p> <p><i>Étude cas-témoin</i> – Rapporter le nombre de sujets dans chaque catégorie d'exposition, ou les indicateurs du niveau d'exposition mesurés</p> <p><i>Études transversales</i> - Rapporter le nombre d'évènements ou les indicateurs mesurés</p>			
Résultats Principaux	16	<p>(a) Indiquer les estimations non ajustées et, le cas échéant, les estimations ajustées pour les facteurs confondants et leur précision (par exemple, intervalle de confiance de 95%). Indiquer clairement quels facteurs confondants ont été ajustés et la raison pour laquelle ils ont été inclus</p>			

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

		(b) Mentionner les limites des intervalles quand des variables continues ont été catégorisées (c) Le cas échéant, envisager de traduire les estimations du risque relatif en risque absolu pour une période de temps significative			
Autres analyses	17	Indiquer les autres analyses faites - par exemple, analyses de sous-groupes et d'interactions, et analyses de sensibilité			
Discussion					
Résultats principaux	18	Résumer les résultats principaux en faisant référence aux objectifs de l'étude			
Limites	19	Discuter des limites de l'étude, en tenant compte des sources de biais potentiels ou d'imprécisions. Discuter de la tendance et de l'ampleur de tout biais potentiel		RECORD 19.1: Discuter les implications de l'utilisation de données qui n'ont pas été créées ou collectées pour répondre aux questions de recherche spécifiques. Inclure une discussion sur les biais de classification erronée, les facteurs confondants non mesurés, les données manquantes et l'évolution de l'admissibilité au fil du temps, en ce qui a trait à l'étude faisant l'objet du rapport.	
Interprétation	20	Fournir une interprétation globale prudente des résultats compte tenu des objectifs, des limites de l'étude, de la			

		multiplicité des analyses, des résultats d'études similaires, et tout autre élément pertinent			
Généralisabilité	21	Discuter de la généralisabilité (validité externe) des résultats de l'étude			
Autres Information					
Financement	22	Indiquer la source de financement et le rôle des financeurs dans la présente étude et, le cas échéant, dans l'étude originale sur laquelle est fondé le présent article			
Accessibilité du protocole, des données brutes et du code de programmation		..		RECORD 22.1: Les auteurs doivent fournir des informations sur la façon d'accéder à toute information supplémentaire telle que le protocole d'étude, les données brutes ou le code de programmation.	

*Référence: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015;12(10): e1001885.

* La liste de contrôle est protégée par la licence Creative Commons Attribution (CC BY).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Confidential

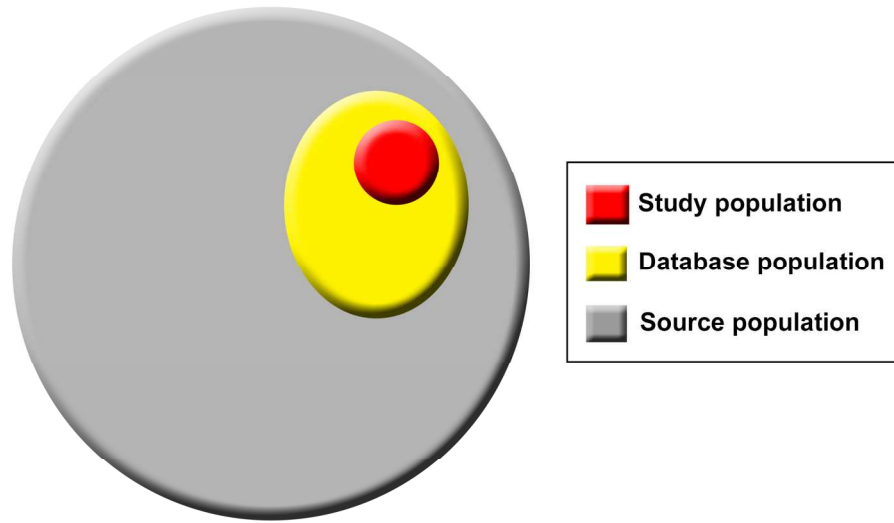


Figure 1. Hiérarchie des populations dans les études menées en utilisant des sources de données collectées de manière routinière.

190x108mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

UNIVERSITÉ SAINT-JOSEPH DE BEYROUTH

FACULTÉ DES LANGUES
ÉCOLE DE TRADUCTEURS ET D'INTERPRÈTES
DE BEYROUTH

Licence en traduction
والإجازة في الترجمة

جامعة القديس يوسف في بيروت
كليه اللغات
مدرسة الترجمة بيروت

إبت زبينة جليلية القسيس أوبت في بيروت
رغم تلبية اللغات
وتعتبر مدرسة الترجمة بيروت
بناء على قرار لجنة الاعتمادات
الترجمة في ١٢ تموز ٢٠١٧

منذ
يارا يوسف بطرس
البروت (لبنان)
١٩ آذار ١٩٩٦
شارم

والإجازة في الترجمة
لغة ا الفرنسية - اللغة العربية - الإنجليزية
لغة ا العربية - اللغة الفرنسية
بيروت في ٢٥ تموز ٢٠١٧

12 juillet 2017
Yara Youssef BOUTROS
Beyrouth (Liban)
19 mars 1996

La licence en traduction
Langue A : Arabe - Langue B : Français - Langue C : Anglais

Date d'épreuve de 25 juillet 2017

٢٠١٧/٦/٢٥

248x361mm (72 x 72 DPI)