

l'exactitude

l'actualité

la comparabilité

la facilité  
d'utilisation

la pertinence

## Le cadre de la qualité des données de l'ICIS

2009



Institut canadien  
d'information sur la santé

Canadian Institute  
for Health Information

Tous droits réservés.

Aucune section de la présente publication ne peut être reproduite ni transmise sous quelque forme que ce soit, ni par quelque procédé électronique ou mécanique que ce soit, y compris les photocopies et les enregistrements, ou par tout autre moyen de stockage d'information et de recherche documentaire, qui existe ou non à l'heure actuelle, sans le consentement écrit de l'auteur qui en possède les droits. Ne sont cependant pas soumises à cette règle les personnes désirant citer de brefs passages pour une publication dans une revue, un journal ou encore pour une émission radiodiffusée ou télédiffusée.

Les demandes d'autorisation doivent être acheminées à :

Institut canadien d'information sur la santé  
495, chemin Richmond, bureau 600  
Ottawa (Ontario) K2A 4H6

Téléphone : 613-241-7860  
Télécopieur : 613-241-8120  
[www.icis.ca](http://www.icis.ca)

ISBN 978-1-55465-697-4 (PDF)

© 2010 Institut canadien d'information sur la santé

Comment citer ce document :

Institut canadien d'information sur la santé, *Le cadre de la qualité des données de l'ICIS 2009*, Ottawa (Ont.), ICIS, 2009.

This publication is also available in English under the title: *The CIHI Data Quality Framework, 2009*.  
ISBN 978-1-55465-696-7 (PDF)

# Le cadre de la qualité des données de l'ICIS

2009

## Table des matières

Remerciements.....	iii
1 Introduction .....	1
1.1 Qu'est-ce que le cadre de la qualité des données de l'ICIS?.....	1
1.2 Quoi de neuf dans la version 2009 du cadre de la qualité des données? .....	2
1.3 Qu'est-ce que la qualité des données et de l'information? .....	2
2 Cycle de travail de la qualité des données .....	4
3 Évaluation de la qualité des données .....	6
3.1 L'outil d'évaluation .....	6
3.2 Dimensions de la qualité des données .....	7
3.3 Intégration des dimensions de la qualité des données .....	8
3.4 Cotes.....	8
3.5 Plan d'action.....	9
4 Documentation sur la qualité des données et de l'information .....	11
4.1 Rapport d'évaluation de la qualité des données.....	11
4.2 Documentation sur la qualité des données à l'intention des utilisateurs .....	13
4.3 Documentation sur les métadonnées.....	15
Annexe A — Changements importants par rapport à la version 2005 .....	17
Annexe B — Sommaire des dimensions, des caractéristiques et des critères.....	19
Annexe C — Outil d'évaluation du cadre de la qualité des données .....	23
Annexe D — Modèle de rapport d'évaluation de la qualité des données.....	87
Annexe E — Sous-catégories pour la documentation sur les métadonnées.....	113
Annexe F — Glossaire .....	125
Bibliographie .....	136



## **Remerciements**

L'Institut canadien d'information sur la santé (ICIS) souhaite remercier les nombreuses personnes, tant à l'ICIS qu'à Statistique Canada, qui ont contribué à l'élaboration de la présente version ainsi qu'aux versions précédentes du cadre de la qualité des données de l'ICIS.



# 1 Introduction

La qualité des données et de l'information fait partie intégrante du mandat de l'Institut canadien d'information sur la santé (ICIS), qui consiste à éclairer les politiques publiques, à appuyer la gestion des soins de santé et à sensibiliser le public aux facteurs qui influent sur la santé. L'ICIS déploie des efforts rigoureux pour s'assurer que les données recueillies et fournies respectent les normes les plus élevées.

L'ICIS a mis en œuvre un programme complet qui comporte des processus et des politiques visant à améliorer continuellement la qualité des données et de l'information au sein de l'ICIS et du secteur de la santé en général. La stratégie de l'ICIS, comprise dans le plan d'action à six volets ci-dessous, inclut un certain nombre d'initiatives de prévention, de dépistage précoce et de résolution de problèmes liés à la qualité :

- encourager le développement d'une culture de la qualité des données;
- renforcer l'infrastructure et la capacité de l'ICIS en matière de qualité des données;
- enrichir la chaîne d'approvisionnement des données;
- accroître la collaboration externe;
- favoriser la communication et fournir des consultations;
- établir un fonds réservé aux projets prioritaires.

Pour connaître la stratégie intégrale de l'ICIS en matière de qualité des données, consultez les documents *Gagner la confiance* et *Gagner la confiance — bilan des trois dernières années*, accessibles sur le site [www.icis.ca](http://www.icis.ca).

L'amélioration de la qualité des données et de l'information relève d'un effort de collaboration entre le personnel de l'ICIS et les fournisseurs de données ainsi que les utilisateurs des données de l'ICIS. Cette collaboration est nécessaire pour satisfaire aux exigences et aux attentes changeantes et grandissantes des utilisateurs des banques de données, qui ont incité l'ICIS à réviser sur une base régulière son cadre de la qualité des données.

## 1.1 Qu'est-ce que le cadre de la qualité des données de l'ICIS?

Le cadre de la qualité des données offre une approche objective visant à mettre en œuvre des processus uniformes de cheminement des données axés sur les priorités en matière de qualité, à évaluer la qualité des données d'une banque de données et à produire de la documentation standard liée à cette banque dans le but ultime d'améliorer continuellement la qualité des données des banques de l'ICIS. Le cadre englobe trois composantes principales :

- le cycle de travail de la qualité des données;
- l'outil d'évaluation de la qualité des données;
- la documentation sur la qualité des données.

**NOUVEAU**

## **1.2 Quoi de neuf dans la version 2009 du cadre de la qualité des données?**

La version 2009 du cadre de la qualité des données de l'ICIS est une mise à jour de la version 2005. Au cours des dernières années, les types de banques de données de l'ICIS se sont considérablement diversifiés. En raison de cette croissance, il était clair que le cadre devait être amélioré par rapport à la version 2005. Cette amélioration était surtout nécessaire en raison du caractère unique des banques de données sur le personnel de la santé, les médicaments, les dépenses de santé, l'équipement médical, les services à domicile et les soins de longue durée par rapport aux banques de données cliniques, sur lesquelles s'appuyaient en grande partie les versions précédentes du cadre de la qualité des données.

### **Les difficultés de l'évaluation des banques de données de l'ICIS**

Étant donné la diversité des banques de données de l'ICIS, l'évaluation de l'ensemble de ces banques à l'aide d'un seul outil n'est pas une mince tâche. Au nombre des difficultés que cela suppose figurent les suivantes :

1. plusieurs niveaux de fournisseurs de données par opposition à un seul fournisseur de données;
2. données agrégées par opposition aux données au niveau de l'enregistrement;
3. soumission volontaire par opposition à la soumission obligatoire;
4. données d'enquêtes par sondage par opposition aux données des recensements;
5. données longitudinales par opposition aux données ponctuelles.

La toute dernière version du cadre compose avec ces difficultés de la façon la plus complète possible. L'annexe A décrit les changements apportés par rapport à la version de 2005.

## **1.3 Qu'est-ce que la qualité des données et de l'information?**

À l'ICIS, la définition de la qualité des données est fonction des utilisateurs des données. Parmi ceux-ci figurent les planificateurs des systèmes de santé, les ministères de la Santé, les autorités régionales, les fournisseurs de données, les associations professionnelles de la santé, les chercheurs, les dispensateurs de soins de santé et d'autres organismes de santé d'intérêt particulier. Si les données de l'ICIS répondent aux besoins précis de ces utilisateurs, on dit alors qu'elles sont adaptées à leur utilisation. Selon le U.S. Census Bureau en 2006, pour être adaptées à leur utilisation, les données doivent posséder les trois attributs suivants : utilité, objectivité et intégrité. L'utilité se rapporte au caractère pertinent des données ou de l'information du point de vue des utilisateurs. L'objectivité indique si les données ou l'information sont exactes, fiables et sans biais, et si elles sont présentées de façon exacte, claire et sans biais. Enfin, l'intégrité se rapporte à la sécurité ou à la protection des données ou de l'information contre les révisions et les accès non autorisés. Les décisions concernant la gestion du système de soins de santé s'appuient sur l'information que fournissent les données. Les données représentent un atout pour toute organisation. Elles sont à la base de la production d'information à partir de laquelle les

décisions sont prises. La gestion de la qualité de l'information au sein d'un collecteur secondaire de données comme l'ICIS revêt donc une importance capitale. Pour parvenir à de l'information de qualité supérieure, il faut déterminer les causes fondamentales des anomalies, prévenir les erreurs dans les processus d'information, établir les exigences en matière de qualité de l'information et contrôler les processus d'information. En appliquant ces principes solides de gestion à l'information, l'ICIS conservera sa réputation de gardien de confiance pour les données pancanadiennes de qualité sur la santé.

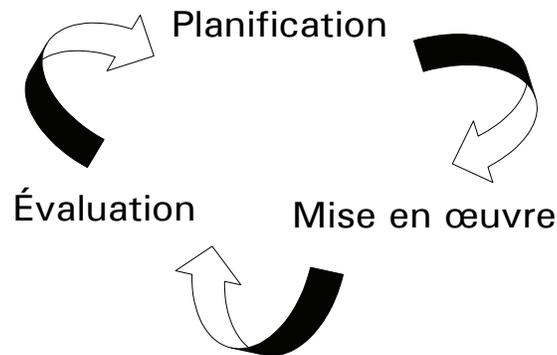
## 2 Cycle de travail de la qualité des données

L'établissement de processus structurés de cheminement des données axés sur les priorités en matière de qualité se fonde sur la définition claire, au sein d'une organisation, des rôles et des responsabilités en matière de qualité des données et sur la mise en œuvre d'un cycle de travail cohérent pour la manipulation des données et de l'information.

À l'ICIS, la qualité des données est la responsabilité de chaque membre du personnel, sans égard au poste occupé ou au service. La qualité des données et de l'information est une priorité de l'organisme. La haute direction oriente de façon générale la qualité des données et de l'information, établit les priorités dans le cadre des projets internes et des projets à long terme sur la qualité des données et veille à l'apport adéquat en ressources pour la réalisation de ces initiatives. Un conseil de la qualité des données, formé de gestionnaires des banques de données et d'utilisateurs de ces banques, se réunit chaque mois et fait rapport à la haute direction, au besoin. Chaque section de l'ICIS responsable d'une banque de données assume certaines tâches, notamment l'analyse et l'évaluation de leurs données et de leurs processus de données afin de relever des problèmes de qualité, la préparation de plans visant à résoudre ces problèmes, l'élaboration et la mise en œuvre de pratiques pour améliorer la banque de données, la réalisation d'études spéciales et la consignation de la qualité de leurs données à l'intention des utilisateurs internes et externes. L'ICIS a créé une division qui a pour mandat d'aider les utilisateurs et le personnel de chaque section responsable d'une banque de données à trouver les problèmes de qualité des données et à les résoudre. Pour ce faire, cette division a créé des outils tels que le cadre de la qualité des données, offre des conseils de nature méthodologique et statistique, réalise des études spéciales, explore les données, produit des outils de déclaration tels que les rapports provinciaux et territoriaux sur la qualité des données à l'intention des sous-ministres de la Santé et engage une discussion avec les intervenants externes au sujet de la qualité des données et de la façon de la mettre en œuvre dans l'ensemble du système de soins de santé.

Le document *CIHI's Database Support Resource Guide* donne des précisions concernant les rôles et responsabilités de la Division de la qualité des données et des sections de l'ICIS dans la mise en œuvre du présent cadre de la qualité des données. Le personnel de l'ICIS peut y accéder à partir de la page intranet de la Division de la qualité des données.

Le **cycle de travail de la qualité des données** utilisé à l'ICIS comprend trois types d'activités : la **planification** d'une activité liée à la qualité des données (c'est-à-dire la modification d'un ensemble de vérifications appliquées aux données à la source afin d'en assurer la qualité avant qu'elles soient versées dans la base de données), la **mise en œuvre** de cette activité et, enfin, l'**évaluation** de l'activité, soit déterminer si l'objectif visé a été atteint (à savoir si les vérifications ont été écrites correctement et si seules des données de qualité supérieure ont été versées dans la base de données). Ces trois types d'activités sont réitérés aussi souvent qu'il est nécessaire pour que soit atteint l'objectif de qualité des données.



La **planification** comprend les activités nécessaires pour préparer et prioriser les processus requis pour une banque de données, de même que la conception de tout changement nécessaire.

La **mise en œuvre** comprend l'élaboration des processus nécessaires et leur application à la banque de données (comme recueillir des données, surveiller les données reçues et publier des rapports écrits). Les résultats des activités de mise en œuvre relatifs à un processus peuvent être utiles dans la planification future de processus similaires.

L'**évaluation** consiste à évaluer la qualité de la banque de données et à déterminer si des changements aux processus sont requis (comme procéder à une évaluation de la qualité des données, rédiger la documentation sur la qualité des données et donner priorité aux améliorations requises). Les modifications qui s'imposent sont apportées à l'étape de la planification. Ainsi, le cycle est itératif et continu. Comme il a été mentionné plus haut, les résultats des activités d'évaluation relatifs à un processus peuvent contribuer à la réalisation des activités de planification futures.

Le cycle de travail en trois étapes est flexible et peut donc être appliqué à n'importe quelle phase d'élaboration et de mise à jour continue d'une banque de données. En misant sur l'intégration de processus axés sur la qualité au sein de chaque activité du cycle de travail, on s'assure de l'efficacité et de la qualité du produit.

## 3 Évaluation de la qualité des données

L'évaluation de la qualité des données de toutes les banques de l'ICIS se fait au moyen de l'**outil d'évaluation**, la composante de base du cadre. L'outil examine chaque aspect de la qualité des données. Cependant, cette dernière étant très vaste, aucun outil unique ne peut l'évaluer de façon exhaustive. L'outil d'évaluation cerne les forces des données et des pratiques relatives à la banque de données évaluée, ainsi que les problèmes à résoudre. Un plan d'action est requis pour déterminer comment et quand ces problèmes seront réglés.

L'outil d'évaluation examine cinq aspects, ou dimensions, de la qualité des données d'une banque de données : l'exactitude, l'actualité, la comparabilité, la facilité d'utilisation et la pertinence. La section 3.2 définit en détail ces cinq dimensions. Le fait de diviser la qualité des données en composantes permet à l'utilisateur de déterminer rapidement les aspects susceptibles de nuire à l'adaptation des données à leur utilisation. Même si bon nombre d'organismes structurent la qualité des données en dimensions très similaires, il n'existe pas de description unique pour ces dimensions. En outre, bien que les cinq dimensions de l'ICIS ne portent pas nécessairement le même nom que celles du cadre de Statistique Canada, elles les englobent toutes. Les organismes nomment peut-être ces dimensions différemment, mais il reste que l'exploration des divers aspects de la qualité des données est une pratique reconnue.

Le but principal de l'outil d'évaluation consiste à évaluer et à documenter les limites et les forces actuelles des banques de données de l'ICIS. Habituellement, les limites sont signalées par des critères avec la cote *non respecté*. Une fois ces problèmes connus, ils peuvent servir à documenter l'état actuel et à recommander des améliorations. L'outil d'évaluation saisit aussi les forces d'une base de données et permet ainsi de dégager les meilleures pratiques de l'ICIS et de les partager entre les sections.

### 3.1 L'outil d'évaluation

L'outil d'évaluation comprend des dimensions, des caractéristiques et des critères. Chaque dimension est composée de caractéristiques connexes et chaque caractéristique se compose de plusieurs critères. Ces derniers consistent en des questions portant sur chaque aspect de la qualité des données. Ce sont les réponses à ces questions qui informent la section de l'ICIS de la présence de problèmes de qualité des données.

**Dimensions (5)**



**Caractéristiques (19)**



**Critères (61)**

Le cadre de la qualité des données de l'ICIS se fonde sur 5 dimensions de la qualité des données, 19 caractéristiques et 61 critères. Chaque critère consiste en une question visant un aspect des données ou de la banque de données. La cote *respecté* ou *non respecté* est généralement accordée à chaque critère. Dans certains cas, les critères sont cotés en fonction d'une fourchette de valeurs. La réponse à chaque question guide le personnel de la banque de données dans la création du plan d'action décrit à la section 3.5.

L'annexe B présente un résumé de cet outil d'évaluation et l'annexe C, l'outil d'évaluation intégral du cadre de la qualité des données.

## 3.2 Dimensions de la qualité des données

Les dimensions sont des composantes distinctes de la définition plus vaste de la qualité des données. Les cinq dimensions de la qualité utilisées à l'ICIS sont définies ci-dessous.

### 1. Exactitude

La dimension de l'exactitude porte sur la conformité de l'information contenue dans la banque de données, ou qui en découle, à la réalité qu'elle doit mesurer.

### 2. Actualité

L'actualité désigne principalement le caractère courant ou à jour des données au moment de leur diffusion selon l'écart entre la fin de la période de référence à laquelle les données se rapportent et la date à laquelle les données deviennent accessibles aux utilisateurs.

### 3. Comparabilité

La dimension de la comparabilité évalue le degré de cohérence des bases de données au fil du temps et leur utilisation des conventions standard (comme des éléments de données ou des périodes de référence), qui rendent ces bases comparables à d'autres bases de données.

### 4. Facilité d'utilisation

La facilité d'utilisation désigne la facilité avec laquelle on peut comprendre les données d'une banque de données et y accéder.

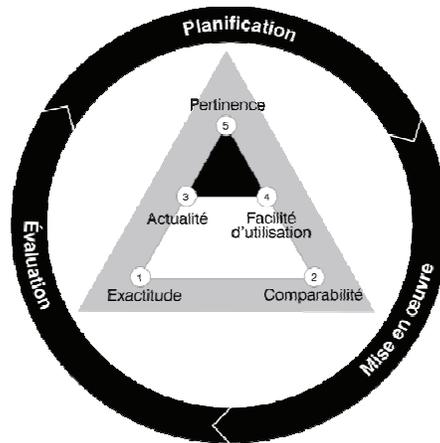
### 5. Pertinence

La pertinence décrit de quelle façon une banque de données répond aux besoins actuels et futurs des utilisateurs.

NOUVEAU

### 3.3 Intégration des dimensions de la qualité des données

Les dimensions peuvent se chevaucher et être étroitement liées. L'évaluation de l'adaptation des données à leur utilisation suppose une gestion adéquate de chaque dimension. En outre, la non-conformité d'une dimension peut compromettre l'utilité du rapport final ou de la publication de données. Un bon équilibre quant au nombre de problèmes décelés pour chaque dimension est donc primordial. Les mesures touchant une dimension de la qualité peuvent influencer favorablement sur les autres dimensions. Le diagramme hiérarchique ci-dessous illustre les cinq dimensions au sein du cycle de travail de la qualité des données décrit à la section 2.



Comme l'indique le diagramme, la dimension la plus importante est la pertinence. En effet, les données qui ne sont pas pertinentes perdent considérablement de la valeur, même si les quatre autres dimensions sont respectées. Les données pertinentes sont celles qui répondent aux besoins actuels et futurs des utilisateurs. Selon les bases de ce modèle hiérarchique, les données doivent être exactes et reposer sur des conventions cohérentes pour être comparables à des données similaires. Une fois ces bases établies, les données doivent être le plus à jour possible pour que les décisions soient fondées sur de l'information récente. Enfin, la facilité d'utilisation se traduit par des données accessibles et faciles à comprendre. Ces quatre dimensions de la qualité des données, si elles sont respectées, permettent à l'utilisateur de prendre des décisions fondées sur des données exactes, comparables, actuelles et faciles à utiliser.

### 3.4 Cotes

L'outil d'évaluation fait appel à des cotes pour souligner les forces et cerner les limites des données. Les cotes se veulent un guide sur lequel s'appuient les sections pour juger de l'adaptation des données à leur utilisation. Aucun système de cotes n'est capable de reconnaître tous les problèmes de qualité des données en raison de la nature subjective de cette dernière et des objectifs différents des banques de données. Des cotes élevées n'indiquent pas forcément qu'une banque de données ne présente pas de problème. De la même façon, des résultats qui signalent des problèmes de qualité des données ne signifient pas nécessairement que les données d'une banque ne doivent pas être utilisées. Il revient à la section de déterminer s'il y a présence ou non de problèmes de qualité dans leurs banques de données, ainsi que de documenter les résultats de l'évaluation afin d'en informer les utilisateurs.

Dans la plupart des cas, une des cotes suivantes est attribuée à chaque **critère** de l'outil d'évaluation : *respecté, non respecté, inconnu* ou *non applicable*. Dans certains cas, les critères reçoivent la cote *minimal ou aucun, modéré, significatif* ou *inconnu* en fonction d'autres catégories prédéterminées. Chaque critère comporte un énoncé ou un tableau de cotes servant à déterminer la cote à assigner. Les sections sont chargées d'attribuer les cotes à leurs banques de données. Les cotes facilitent l'établissement de comparaisons normalisées entre les bases de données et entre les différentes années de données au sein d'une même banque.

**Il convient de souligner que les cotes sont réservées à un usage interne et ne figurent que dans le rapport d'évaluation. Elles doivent être exclues de tout document externe.**

NOUVEAU

### 3.5 Plan d'action

Une fois l'évaluation de la banque de données terminée, les principales conclusions sont résumées et un plan d'action des stratégies qui serviront à corriger les lacunes est dressé. Le tableau A donne un exemple de plan d'action qui précise les recommandations, le calendrier de mise en œuvre et la ou les personnes responsables. Les gestionnaires de sections doivent effectuer un suivi afin de s'assurer que les améliorations sont effectivement apportées.

**Tableau A Exemple de plan d'action**

Recommandation	Date de mise en œuvre	Responsables	Date butoir ou processus continu	Participation de la Division de la qualité des données
1. Rechercher la cause des doubles contribuant au surdénombrement	2007-2008	Alain	2008-2009	Oui
2. Évaluer les éléments de données en fonction du Dictionnaire de données de l'ICIS	2008-2009	Barbara	Septembre 2008	Non
3. Effectuer des contrôles de validité pour chacun des éléments de données et signaler les données invalides	2006-2007	Barbara Caroline	Annuellement	Non

<b>Recommandation</b>	<b>Date de mise en œuvre</b>	<b>Responsables</b>	<b>Date butoir ou processus continu</b>	<b>Participation de la Division de la qualité des données</b>
4. Créer la documentation principale des méthodes	2007-2008	Alain Caroline Donald	2008-2009	Oui
5. Évaluer la satisfaction des utilisateurs	2006-2007	Pierre	2007-2008 Décembre 2008 (nouvelle date butoir)	Non

## 4 Documentation sur la qualité des données et de l'information

La **documentation sur la qualité des données et de l'information** fait partie intégrante du cadre de la qualité des données de l'ICIS. Le fait d'informer les utilisateurs, tant à l'interne qu'à l'externe, de la qualité des données de l'ICIS occupe une place importante dans le programme de la qualité des données de l'ICIS, car cela leur permet de décider si les données conviennent ou non à l'utilisation qu'ils désirent en faire.

La documentation est un volet essentiel de toute banque de données. Elle fournit le contexte approprié pour l'information, une confirmation facile des faits et, entre autres, une méthode efficace de diffusion et de stockage de l'information. Bien que de nombreux types de documentation soient nécessaires pour une base de données, le présent chapitre met l'accent sur la documentation liée à la qualité des données uniquement.

Le cadre de la qualité des données de l'ICIS exige trois types de documentation relative à la qualité des données d'une banque :

1. *Rapport d'évaluation de la qualité des données* : rapport interne de l'ICIS qui résume les résultats de l'évaluation de la qualité des données;
2. *Documentation sur la qualité des données à l'intention des utilisateurs* : documentation fournie aux utilisateurs de la banque de données;
3. *Documentation sur les métadonnées* : documentation détaillée sur la banque de données.

### 4.1 Rapport d'évaluation de la qualité des données

#### 4.1.1 Objectif

Le rapport d'évaluation de la qualité des données s'adresse principalement au personnel de la banque de données. Ce rapport interne fait le compte rendu des résultats de l'évaluation réalisée en vue de cerner les forces de la banque de données ainsi que tout problème de qualité des données. La documentation sur la qualité des données à l'intention des utilisateurs fait l'objet d'une publication externe et son contenu repose largement sur le rapport interne.

Consultez la section 4.1.4 pour connaître le contenu et la présentation proposés de ce rapport.

#### 4.1.2 Quand produire ce rapport

L'évaluation devrait porter sur un sous-ensemble de données récent. Il peut s'agir de l'année la plus récente si les données sont recueillies sur une base annuelle, ou d'une version récente de la base de données complète. Les cotes doivent refléter l'état actuel de la banque de données, mais les changements ou les améliorations à venir peuvent tout de même être indiqués. Il est possible de répondre à certaines questions liées aux critères sans attendre que toutes les données soient reçues. Pour les autres questions, toutes les données sont nécessaires avant que le rapport final d'évaluation puisse être réalisé. Chaque

critère doit être évalué avant toute publication majeure de données ou la publication d'un rapport annuel. Le rapport d'évaluation interne doit être produit au plus tard trois mois après la diffusion des données ou d'une publication.

### 4.1.3 Examen et soutien du personnel de la qualité des données

Le personnel de la Division de la qualité des données peut donner des conseils lors du processus d'évaluation des critères et offrir son soutien dans l'élaboration du rapport et la révision de la version finale. Plus précisément, une fois le rapport rédigé par le personnel de la banque de données, le personnel de la Division de la qualité des données le révise et donne ses commentaires durant le mois en cours. Une fois les commentaires pris en compte, le gestionnaire de la banque de données donne son approbation au rapport d'évaluation final. L'approbation donne l'aval aux activités requises pour améliorer la qualité des données précisées dans le plan d'action du rapport d'évaluation. Lorsque le rapport d'évaluation de la qualité des données a été approuvé par la direction, un exemplaire est envoyé à la Division de la qualité des données, puis le rapport est publié sur la page intranet de la Division.

**NOUVEAU**

Outre cette révision régulière du rapport et le soutien du personnel de la qualité des données, une analyse critique d'une banque de données sélectionnée et du plus récent rapport d'évaluation de la qualité des données est réalisée chaque mois. Quelques membres du personnel de la Division de la qualité des données ainsi que deux évaluateurs externes (habituellement des gestionnaires d'autres sections) rencontrent le gestionnaire et l'équipe de la section dont le rapport fait l'objet de l'analyse critique. Les personnes présentes discutent simplement de différentes questions comme de l'utilisation adéquate de l'outil d'évaluation et des pratiques qui permettraient d'améliorer le cadre de la qualité des données. Afin de favoriser la communication des pratiques et des outils axés sur la qualité des données au sein de l'organisation, le personnel de la banque de données sélectionnée présente, à la suite de l'analyse critique, les points saillants soulevés lors de la plus récente utilisation de l'outil d'évaluation dans le cadre d'un forum interne.

### 4.1.4 Contenu et présentation proposés

Le rapport d'évaluation doit contenir suffisamment d'information concernant chaque critère pour justifier la cote. Les personnes qui ne connaissent pas bien la banque de données sont ainsi à même de comprendre pourquoi cette cote a été attribuée. De façon similaire, les critères auxquels la cote *non applicable* a été attribuée doivent être accompagnés d'une remarque expliquant pourquoi ils ne s'appliquent pas.

Le rapport devrait commencer par une brève introduction qui présente l'objectif de la banque de données, l'utilisation qui en est faite, une description et le délai d'exécution de l'évaluation. Les critères doivent être évalués selon l'ordre où ils figurent dans l'outil d'évaluation. Les recommandations sont une composante importante du processus d'évaluation. Elles devraient donc être intégrées au rapport même en regard des critères ou des caractéristiques problématiques. Elles doivent aussi figurer dans le sommaire. L'annexe D contient une version Word d'un modèle de rapport d'évaluation.

## **4.2 Documentation sur la qualité des données à l'intention des utilisateurs**

### **4.2.1 Objectif**

Le but de la documentation sur la qualité des données à l'intention des utilisateurs est de fournir à ces derniers suffisamment d'information pour leur permettre de déterminer si les données sont adaptées à leur utilisation. Il est important de présenter les problèmes de qualité des données en fonction du contexte d'utilisation et de l'incidence qu'ils peuvent avoir sur l'analyse de l'utilisateur. Bien que cela ne soit pas obligatoire, la documentation sur la qualité des données peut être intégrée à un rapport annuel selon la fréquence de présentation des résultats et les types de résultats présentés. Dans le cas de certaines banques de données pour lesquelles un rapport annuel est produit, l'intégration de la documentation sur la qualité des données à la section du rapport traitant des notes méthodologiques ou des limites des données répond aux besoins des utilisateurs. Toutefois, dans le cas d'autres banques de données, il est recommandé de créer un document indépendant et complet de la qualité des données renfermant tous les renseignements pertinents dont les utilisateurs ont besoin. Pour toute banque de données associée à de nombreuses diffusions (sous forme de données ou de rapports), il serait judicieux d'envisager un document indépendant sur la qualité des données afin que les limites des données soient systématiquement mises en évidence, facilement accessibles et fassent l'objet d'une publication distincte.

### **4.2.2 Contenu et présentation proposés**

La documentation sur la qualité des données à l'intention des utilisateurs d'une banque doit permettre à ces derniers de bien comprendre l'ensemble des limites des données. Le personnel de la Division de la qualité des données offre son soutien au moment de l'élaboration du rapport et révisé la version finale.

La section qui suit précise le contenu proposé. Chaque banque de données peut décider de la façon de présenter son document sur les limites des données. Toutefois, chacun des sujets suivants doit être abordé.

#### **i) Introduction**

L'introduction devrait indiquer l'objectif du document de l'ICIS sur la qualité des données, notamment à l'aide d'un ou deux paragraphes décrivant les données évaluées, leur raison d'être, un résumé des principales limites des données ainsi que des références fournissant plus d'information sur la source des données. L'introduction devrait aussi comprendre des renseignements détaillés sur l'intervalle de temps couvert par le document, dont les dates précises d'inclusion ou d'exclusion.

## **ii) Concepts et définitions**

### **a) Raison d'être ou objectif**

La raison d'être ou l'objectif de la banque de données doit être précisé.

### **b) Population**

La population de la banque de données, c'est-à-dire habituellement la population de référence, devrait être décrite et une période précise, ajoutée.

### **c) Éléments de données et concepts**

Les principaux éléments de données et concepts doivent être définis. Il n'est pas nécessaire de fournir des formules exactes ni de reproduire le contenu du Dictionnaire de données de l'ICIS. Seule l'information pertinente est essentielle.

## **iii) Principales limites des données**

Cette section a pour objectif d'informer les utilisateurs des principaux problèmes de qualité des données. Elle devrait porter sur les principales limites des données aussi bien que sur leur incidence et les solutions prévues, et discuter de toutes les parties de la banque de données problématiques selon l'outil d'évaluation de la qualité des données.

Bien qu'il soit possible d'inclure toutes les limites d'une source de données, le document sur la qualité des données à l'intention des utilisateurs peut se restreindre aux problèmes et aux variables qui intéressent les principaux utilisateurs. Si une demande d'un utilisateur externe ou un produit précis exige un document distinct sur la qualité des données, seules les limites pertinentes doivent être abordées.

## **iv) Couverture**

Cette section devrait comporter trois parties : une description de la base de sondage, une description des procédures de mise à jour de la base de sondage et une description de l'incidence de ces procédures.

## **v) Collecte et non-réponse**

Les descriptions de cette section devraient être aussi brèves que possible. Bien qu'il puisse être très tentant de décrire en détail tout le travail effectué pour la collecte et l'assemblage des données, tel n'est pas le but de cette section. Elle vise plutôt à expliquer comment les procédures de collecte, entre autres, peuvent influencer sur l'analyse des utilisateurs externes. Elle devrait traiter de la collecte des données, des procédures de contrôle de la qualité, des taux de réponse applicables, de tous les ajustements à la suite de la non-réponse (ou absence d'ajustements) et de tout autre problème connu de biais ou de fiabilité.

#### **vi) Modifications méthodologiques d'envergure au cours des ans**

Cette section illustre les changements apportés à la banque de données actuelle par rapport aux versions précédentes (p. ex. les cas où les différences dans les chiffres pourraient être attribuables à une modification des procédures et non pas refléter des changements réels). Cette section devrait inclure toute modification d'envergure apportée à la banque de données dans l'année précédente ainsi que celles qui sont prévues. Si des comparaisons longitudinales sont réalisées, il importe de présenter tout changement antérieur pertinent et de fournir les références à la documentation pertinente.

#### **vii) Historique des révisions**

Cette section présente les modifications ou les corrections apportées aux données qui ont été présentées antérieurement (données rétrospectives). Pour éviter toute confusion lorsque les chiffres révisés ne correspondent plus aux données diffusées antérieurement, il est nécessaire d'informer les utilisateurs de toute modification de quelque nature que ce soit apportée aux données ou aux estimations utilisées dans les années antérieures. Il faut également préciser l'incidence ou la solution prévue de chaque problème majeur.

#### **viii) Comparabilité externe**

La dernière section traite de la comparabilité des données ou des estimations avec d'autres sources. Si les estimations ou les données ne sont pas comparables à des estimations ou à des données similaires qui ont été diffusées antérieurement ou qui le seront par des sources externes à la base de données, il importe d'indiquer les écarts en question.

### **4.3 Documentation sur les métadonnées**

Le troisième type de documentation essentielle à toute banque de données est celui consacré aux métadonnées. Une documentation efficace des processus méthodologiques repose sur la consignation des activités de collecte, de traitement et de diffusion des données, des limites des données et des mesures de la qualité des données au sein *d'un seul document, complet et à jour*. L'ICIS a créé un dépôt de métadonnées qui héberge la documentation sur la qualité des données des banques ainsi que d'autres précieuses métadonnées.

**NOUVEAU**

Le dépôt de métadonnées devrait être la première source consultée pour répondre à des questions précises sur une banque de données. Un des objectifs de la documentation sur les métadonnées est de décrire le cheminement des données d'une banque de façon à ce qu'un nouveau venu dans l'équipe de projet puisse prendre part à n'importe quelle activité. L'information fournie devrait inclure les étapes du traitement des données, les responsables, les personnes chargées d'approuver les étapes et les mesures de vérification de la qualité des données en place. La documentation sur les politiques et les procédures d'une banque de données favorise les activités d'amélioration des processus, les listes de contrôle et la description des responsabilités, et augmente la fiabilité des données. Grâce à une bonne documentation, la banque de données devient davantage tributaire d'un processus que d'une personne, ce qui est susceptible d'améliorer la qualité et la stabilité des données.

La documentation relative au dépôt de métadonnées devrait comprendre les sept catégories de renseignements définies ci-dessous.

1. *Description de la banque de données* : renseignements généraux essentiels sur la banque de données indiquant les sources des données, les utilisations et les utilisateurs des données ainsi que les responsabilités des personnes qui participent à la soumission, au traitement, à l'analyse et à la diffusion.
2. *Méthodologie* : précisions sur la population cible, la population de référence, la sélection de l'échantillon (dans le cas d'un sondage) et la base de sondage, et renseignements sur la mise en œuvre et la consignation des changements apportés à la base de sondage.
3. *Collecte et saisie des données* : procédures ou spécifications relatives aux données recueillies ou soumises et formatées ou normalisées aux fins de traitement.
4. *Traitement des données* : procédures de traitement des données à partir du moment où les données sont saisies par l'ICIS ou à l'ICIS, jusqu'à la production du fichier analytique, par exemple tous les processus de vérification, de dérivation et d'estimation.
5. *Analyse et diffusion des données* : procédures d'analyse et étapes de production ou de diffusion des tableaux, des rapports et des publications, ainsi que réponses aux demandes.
6. *Stockage des données* : stockage des données, résultats ou estimations à partir de la banque de données.
7. *Documentation* : toute documentation pertinente portant sur le niveau de qualité des données atteint ainsi que sur l'amélioration ou le maintien de la qualité des données.

Il est à noter que ces 7 catégories sont divisées en 118 sous-catégories. L'annexe E décrit en détail chacune des sous-catégories comprises dans les catégories de métadonnées.

## Annexe A – Changements importants par rapport à la version 2005

**NOUVEAU**

La section qui suit présente les changements importants intégrés à cette toute nouvelle version du cadre de la qualité des données par rapport à la version 2005. Une comparaison détaillée des changements fera l'objet d'une publication distincte.

1. Le plan d'action, produit à la suite de l'application de l'outil d'évaluation (décrit à la section 3.5), a été remanié de façon à permettre la création d'un outil utile à l'amélioration de la qualité des données des banques grâce à l'assignation de responsabilités pour la mise en œuvre des recommandations formulées.  
La présente version du cadre comprend un modèle de création du plan d'action.
2. Le cadre est maintenant mieux adapté aux banques de données créées à partir d'une enquête réalisée sur un échantillon sélectionné dans la population de référence.
3. Certaines sections ont été mises à jour afin que soient incluses les banques de données pour lesquelles la soumission est volontaire.
4. Certaines sections ont été mises à jour afin que soient incluses les banques de données de nature longitudinale.
5. Le cadre a été modifié de façon à refléter l'existence d'un nouveau dictionnaire de données de l'ICIS ainsi que la finalisation des éléments de données sous la classe d'objets relative aux dispensateurs.
6. La présente version traite de l'élaboration d'une documentation normalisée sur les métadonnées. Un dépôt électronique principal de métadonnées à l'intention du personnel interne de l'ICIS (disponible en anglais seulement sous le nom « Master Metadata e-Repository ») a été mis en place; il héberge l'ensemble de la documentation essentielle sur les métadonnées.
7. Les concepts des sections sur la couverture, la saisie et la collecte, la non-réponse totale, l'erreur de mesure, la vérification et l'imputation, ainsi que le traitement et l'estimation ont été clarifiés. De plus, trois des critères de ces sections ont été divisés en deux parties (partie a et partie b); le nombre total de critères est donc passé de 58 à 61.



## Annexe B – Sommaire des dimensions, des caractéristiques et des critères

Le tableau suivant présente les dimensions de la qualité des données et leurs critères respectifs qui composent le cadre de la qualité des données de l'ICIS de 2009. L'outil d'évaluation du cadre de la qualité des données (annexe C) décrit plus en détail chaque critère des dimensions de la qualité des données ainsi que la façon d'évaluer chacun d'eux aux fins du rapport d'évaluation de la qualité des données.

Dimensions et caractéristiques	Critères
<b>Exactitude</b>	
<b>Couverture</b>	1a La population de référence est explicitement mentionnée dans toutes les diffusions.
	1b Des efforts sont faits pour réduire l'écart entre la population de référence et la population cible.
	2 Les sources connues de sous-dénombrement ou de surdénombrement ont été documentées.
	3 La base de sondage a été validée au moyen d'une comparaison des sources externes et indépendantes.
	4 Le taux de sous-dénombrement ou de surdénombrement se classe parmi l'une des catégories prédéfinies.
<b>Saisie et collecte</b>	5a Les pratiques de l'ICIS qui minimisent le fardeau de réponse sont documentées.
	5b L'ICIS possède de la documentation sur les pratiques des fournisseurs de données qui minimisent le fardeau de réponse.
	6 Il existe des pratiques qui favorisent la collaboration pour la soumission de données.
	7 Il existe des pratiques qui appuient les fournisseurs de données.
	8 Il existe des processus de soumission standard et ils sont suivis par les fournisseurs de données.
	9 Il existe des mesures de contrôle de la qualité de la saisie des données et elles sont mises en œuvre par les fournisseurs de données.

Dimensions et caractéristiques	Critères
<b>Non-réponse totale</b>	10 L'ampleur de la non-réponse totale est mentionnée dans la documentation sur la qualité des données.
	11 Le nombre d'enregistrements reçus pour les unités déclarantes est examiné afin que les valeurs inhabituelles puissent être décelées.
	12 L'ampleur de la non-réponse totale se classe parmi l'une des catégories prédéterminées.
<b>Non-réponse partielle (ou non-réponse par question)</b>	13 La non-réponse partielle est connue.
	14 L'ampleur de la non-réponse partielle se classe parmi l'une des catégories prédéterminées.
<b>Erreur de mesure</b>	15 Le niveau de l'erreur de mesure se classe parmi l'une des catégories prédéterminées.
	16 Le niveau du biais n'est pas significatif.
	17 Le degré de problèmes de cohérence se classe parmi l'une des catégories prédéterminées.
<b>Vérification et imputation</b>	18 Des contrôles de la validité sont réalisés pour chacun des éléments de données et toutes les données invalides sont signalées.
	19 Les règles de vérification et l'imputation sont logiques et appliquées de façon uniforme.
	20 Les rapports de vérification destinés aux utilisateurs sont faciles à utiliser et à comprendre.
	21 Le processus d'imputation est automatisé et conforme aux règles de vérification.
<b>Traitement et estimation</b>	22 La documentation sur toutes les activités de traitement des données est tenue à jour.
	23 Les spécifications techniques pour la banque de données sont tenues à jour.
	24 Les modifications à la structure sous-jacente ou aux programmes de traitement ou d'estimation d'une banque de données ont été testées.
	25 Les données brutes, conformément à la politique de l'ICIS en matière de maintien des données, sont sauvegardées dans un endroit sécuritaire.
	26a Les statistiques agrégées d'une banque de données ont été comparées, dans la mesure du possible, à des statistiques similaires provenant d'une autre banque de données de l'ICIS ou d'une source externe.
	26b La variance de l'estimation, comparée à l'estimation elle-même, se situe à un niveau acceptable.

<b>Dimensions et caractéristiques</b>	<b>Critères</b>
<b>Actualité</b>	
<b>Actualité des données au moment de leur diffusion</b>	27 L'écart entre la date réelle de diffusion et la fin de la période de référence est raisonnablement bref.
	28 La date officielle de diffusion a été annoncée avant la diffusion.
	29 La diffusion a eu lieu à la date officielle prévue.
	30 Les activités de traitement des données sont revues régulièrement aux fins d'amélioration de l'actualité des données.
<b>Actualité de la documentation</b>	31 La documentation recommandée sur la qualité des données était disponible au moment de la diffusion des données ou des rapports.
	32 Les principaux rapports ont été diffusés à la date prévue.
<b>Comparabilité</b>	
<b>Normes du Dictionnaire de données</b>	33 Tous les éléments de données font l'objet d'une évaluation visant à déterminer s'ils sont inclus dans le Dictionnaire de données de l'ICIS.
	34 Les éléments de données d'une banque de données qui sont inclus dans le Dictionnaire de données de l'ICIS doivent être conformes aux normes de ce dictionnaire.
<b>Normalisation</b>	35 Les données sont recueillies au niveau de détail le plus élevé possible.
	36 Pour tout élément de données dérivé, l'élément de données original demeure accessible.
<b>Couplage</b>	37 Les données géographiques sont recueillies au moyen de la Classification géographique type (CGT).
	38 Les données sont recueillies selon le même calendrier, plus particulièrement d'une province ou d'un territoire à l'autre et au sein d'une province ou d'un territoire.
	39 L'utilisation d'identificateurs permet de distinguer les établissements ou les organismes seulement aux fins de couplage rétrospectif.
	40 L'utilisation d'identificateurs permet de distinguer les personnes ou les appareils seulement aux fins de couplage rétrospectif.
<b>Équivalence</b>	41 La méthodologie et les limites des concordances ou des conversions sont documentées.
	42 L'ampleur des problèmes liés aux concordances et aux conversions se classe parmi l'une des catégories prédéterminées.

<b>Dimensions et caractéristiques</b>	<b>Critères</b>
<b>Comparabilité rétrospective</b>	43 La documentation des modifications antérieures apportées à la banque de données existe et est facilement accessible.
	44 La réalisation d'une analyse des tendances permet d'examiner les changements dans les éléments de données de base au fil du temps.
	45 L'ampleur des problèmes découlant de la comparaison des données au fil du temps se classe parmi l'une des catégories prédéterminées.
<b>Facilité d'utilisation</b>	
<b>Accessibilité</b>	46 Un ensemble de données final est rendu disponible à la date prévue de diffusion.
	47 Des tableaux et des analyses au format et au contenu standard sont produits lors de chaque diffusion prévue ou sur demande.
	48 Les produits sont définis, catalogués ou annoncés.
<b>Documentation</b>	49 Il existe de la documentation sur la qualité des données à l'intention des utilisateurs.
	50 Il existe de la documentation sur les métadonnées.
	51 Un avertissement accompagne toute diffusion préliminaire.
<b>Interprétabilité</b>	52 Un mécanisme permet aux principaux utilisateurs de communiquer leurs commentaires à la section d'une banque de données et de recevoir une réponse de cette dernière.
	53 Des lignes directrices de révision sont disponibles et appliquées à chaque diffusion.
<b>Pertinence</b>	
<b>Adaptabilité</b>	54 Des mécanismes tiennent les clients et les intervenants informés des développements dans le domaine.
	55 La banque de données est conçue de manière à faciliter les modifications futures du système.
<b>Valeur</b>	56 Le mandat de la banque de données comble une lacune en matière d'information sur la santé.
	57 Le niveau d'utilisation de la banque de données est surveillé.
	58 La satisfaction des utilisateurs est évaluée périodiquement.

## Annexe C – Outil d'évaluation du cadre de la qualité des données

La présente annexe contient l'outil d'évaluation du cadre de la qualité des données de l'ICIS de 2009, qui décrit les 5 dimensions, les 19 caractéristiques et les 61 critères du cadre de la qualité des données de 2009. Cet outil vise à aider le personnel des banques de données à mieux comprendre le cadre et à effectuer le rapport d'évaluation de la qualité des données connexe.

Il est structuré en sections distinctes, une pour chacune des cinq dimensions de la qualité des données.

1. exactitude
2. actualité
3. comparabilité
4. facilité d'utilisation
5. pertinence

### 1 Dimension de l'exactitude

Pour la plupart des gens, la qualité des données est synonyme d'exactitude. On entend par exactitude la précision avec laquelle l'information contenue dans la banque de données ou qui en découle reflète la réalité qu'elle doit mesurer. Ainsi, le cas échéant, l'information ou les données peuvent être considérées comme exactes ou fiables.

L'exactitude d'une base de données dépend de plusieurs facteurs. Au moment d'examiner l'exactitude, il importe de tenir compte de ce qui suit.

- ✓ Les données appropriées sont-elles toutes présentes?

Trois caractéristiques répondent à cette question.

- La couverture : savez-vous qui devrait soumettre des données?
- La non-réponse totale : les enregistrements ont-ils tous été soumis?
- La non-réponse partielle : les enregistrements soumis sont-ils complets?

- ✓ Les données sont-elles bonnes?

Deux caractéristiques répondent à cette question.

- La saisie et la collecte : quelles sont les mesures utilisées pour minimiser les erreurs dans les données entrantes?
- L'erreur de mesure : les données ont-elles été bien déclarées à l'ICIS?

✓ Que fait-on avec les données?

Deux caractéristiques répondent à cette question.

- La vérification et l'imputation : les contrôles et les modifications aux données sont-ils logiques et cohérents?
- Le traitement et l'estimation : a-t-on mis à l'essai et documenté les processus utilisés pour générer les valeurs?

À partir des questions ci-dessus, la dimension de l'exactitude a été divisée en sept caractéristiques. Celles-ci sont composées de 29 critères évalués de façon individuelle. Il est à noter que trois des critères comportent deux parties (partie a et partie b).

Dimension	Caractéristiques	Critères
Exactitude	Couverture	1 à 4
	Saisie et collecte	5 à 9
	Non-réponse totale	10 à 12
	Non-réponse partielle (ou non-réponse par question)	13 à 14
	Erreur de mesure	15 à 17
	Vérification et imputation	18 à 21
	Traitement et estimation	22 à 26

### Niveaux d'observation



Les calculs relatifs aux trois premières caractéristiques de la dimension de l'exactitude (couverture, saisie et collecte et non-réponse totale) peuvent être réalisés selon plus d'un niveau d'observation : il peut s'agir du niveau des unités de la base de sondage et du niveau au sein de chaque unité de la base de sondage, c'est-à-dire les unités d'analyse. Selon le cas, il est possible qu'une unité de la base de sondage et qu'une unité d'analyse soit la même. Le tableau ci-dessous montre des exemples d'écarts possibles entre les résultats selon le niveau d'observation.

Tableau B Exemples de niveaux d'observation

Ex.	Type de données	Ex.	Fournisseur de données	Niveaux d'observation	Remarques	Unité	Élément
1	Données agrégées	Nombre de diplômés	Facultés de médecine	Unité de la base de sondage (facultés de médecine), unités d'analyse (facultés de médecine)	Identique, puisque les données agrégées sont la somme des enregistrements individuels.  Données agrégées par les facultés de médecine.	Facultés de médecine	Fichier de données agrégées, p. ex. nombre de diplômés
2	Enregistrement individuel	Inscription des infirmières	Associations d'infirmières	Unités de la base de sondage (associations d'infirmières), unités d'analyse (enregistrements individuels des infirmières)	Niveau de l'unité de la base de sondage distinct du niveau de l'unité d'analyse. Évaluations distinctes nécessaires pour les critères 1 à 12.	Associations d'infirmières, inscription des infirmières	Données sur l'inscription des infirmières, p. ex. année d'obtention du diplôme
3	Enregistrement individuel	Dossiers de patients hospitalisés	Ministères de la Santé ou hôpitaux	Unités de la base de sondage (ministères de la Santé ou hôpitaux), unités d'analyse (dossiers des patients)	Deux niveaux d'unités de la base de sondage et un niveau d'unité d'analyse. Évaluations distinctes nécessaires pour les critères 1 à 12.	Ministères de la Santé, hôpitaux, dossiers de patients	Données figurant dans les dossiers des patients, p. ex. sexe et date d'admission

NOUVEAU

Dans les cas où il y a plus d'un niveau d'observation et que l'évaluation vise à déterminer de façon précise la qualité des données et de l'information à chaque niveau, il est essentiel que chacun d'eux soit pris en compte dans l'évaluation. L'évaluation globale d'un critère devrait être fondée sur la pire évaluation des niveaux applicables.

## Couverture

### Critères

- 1a *La population de référence est explicitement mentionnée dans toutes les diffusions.*
- 1b *Des efforts sont faits pour réduire l'écart entre la population de référence et la population cible.*
- 2 *Les sources connues de sous-dénombrement ou de surdénombrement ont été documentées.*
- 3 *La base de sondage a été validée au moyen d'une comparaison des sources externes et indépendantes.*
- 4 *Le taux de sous-dénombrement ou de surdénombrement se classe parmi l'une des catégories prédéfinies.*

Il y a sous-dénombrement ou surdénombrement lorsqu'il y a un écart entre la **population de référence** et la **base de sondage**.

La **population cible** est le groupe d'unités sur lequel on veut obtenir de l'information.

**Par exemple**, la population cible peut être :

*Tous les hôpitaux au Canada ayant au moins un lit de soins de courte durée.*

Toutefois, si l'information relative à la population cible complète n'est pas disponible, il faut y faire référence au moyen d'une population pour laquelle c'est le cas. La **population de référence** représente ce groupe d'unités et il s'agit de celle sur laquelle portent les énoncés.

**Exemple A** : La population de référence pour une base de données peut être :

*Tous les hôpitaux financés par l'État, non situés dans les prisons, qui possédaient au moins un lit de soins de courte durée, dans l'ensemble des provinces et territoires à l'exception de la Saskatchewan et du Nouveau-Brunswick, et qui admettaient des patients au 1<sup>er</sup> janvier de l'année de référence.*

La population de référence d'une banque de données doit être logique et uniforme et, idéalement, le plus près possible de la population cible.

La **base de sondage** d'une banque de données est une liste d'unités (provinces, établissements, médecins, etc.) qui permet d'avoir accès aux unités dans la population de référence faisant l'objet de la collecte. Tel que le précise le tableau B, il est possible qu'une banque de données présente plus d'un niveau d'unité (c'est-à-dire unités de la base de sondage et unités d'analyse). La base de sondage des bases de données administratives, tout comme celle des enquêtes, doit être exacte. **La base de sondage devrait servir à déterminer quelle devrait être la source des données et la proportion de données réellement reçues.**

Il y a **sous-dénombrement** lorsqu'une partie de la population de référence n'est pas incluse dans les unités de la base de sondage. Dans l'exemple A, il y aurait sous-dénombrement si certaines unités d'analyse de la Colombie-Britannique n'étaient pas incluses dans la base de sondage alors qu'elles auraient dû l'être. Il y aurait également sous-dénombrement si une région sanitaire de la Nouvelle-Écosse était divisée en deux, mais que seulement l'une de ces deux régions était incluse dans la base de sondage. Il importe de noter que le sous-dénombrement peut causer une perte d'information (c'est-à-dire des totaux plus faibles) et biaiser les résultats si les unités exclues de la base de sondage diffèrent de celles qui sont incluses.

Il y a **surdénombrement** lorsque des unités ne faisant pas partie de la population de référence sont incluses dans la base de sondage ou lorsque des doubles figurent dans la base de données.

Dans l'exemple A, les données soumises par les établissements ci-dessous entraîneraient un surdénombrement, à moins que des mesures ne soient prises pour corriger les données :

1. un hôpital de la Saskatchewan
2. un hôpital de l'Ontario sans lit de soins de courte durée
3. un hôpital du Québec ayant ouvert ses portes le 1<sup>er</sup> mars

Il y aurait aussi surdénombrement si des données relatives aux mêmes unités étaient fournies à la fois par l'hôpital et par la province; les données seraient donc présentes en double. Le tableau C présente le sommaire des termes définis dans la présente section consacrée à la couverture.

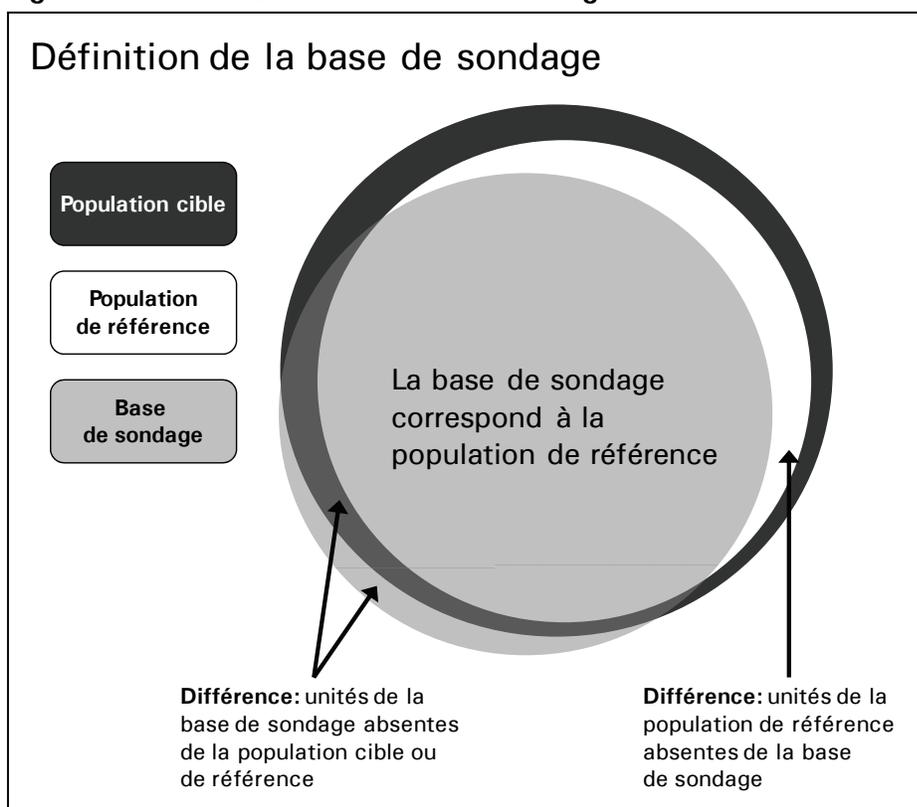
**Tableau C Sommaire des termes liés à la couverture**

NOUVEAU

Terme	Définition
Population cible	Données que l'ICIS souhaite éventuellement recueillir
Population de référence	Données que l'ICIS peut, de façon réaliste, s'attendre à recevoir
Base de sondage	Liste d'unités dans la population de référence à partir de laquelle les données seront recueillies
Surdénombrement	Situation où la base de sondage contient plus de données que prévu (base de sondage > population de référence)
Sous-dénombrement	Situation où la base de sondage contient moins de données que prévu (base de sondage < population de référence)

Afin de clarifier davantage les différences entre population cible, population de référence et base de sondage, la figure 1 représente ces concepts de façon schématisée.

Figure 1 Définition de la base de sondage



Les unités qui ne devraient pas être incluses dans la banque de données et qui ne sont pas incluses dans la **base de sondage** sont considérées comme étant **hors du champ d'observation**.

Comme le recours à une base de sondage adéquate est capital, il est important de la **mettre à jour** périodiquement en y ajoutant de nouvelles unités et en supprimant celles qui ne sont plus dans la population de référence. Toute l'information relative à la base de sondage devrait être tenue à jour. On appelle **procédures de mise à jour de la base de sondage** toute pratique servant à mettre à jour la base de sondage.

Il est important de comprendre que les erreurs de couverture, à l'exception des doubles, ne sont pas nécessairement liées à la soumission des données. Si une des unités de la base de sondage ne soumet pas de données, il s'agit d'un exemple de **non-réponse** et non pas de sous-dénombrement. Dans le cas des bases de données qui utilisent la base de sondage uniquement comme une liste de fournisseurs de données, une vérification externe (p. ex. la consultation des listes de fournisseurs provenant des provinces) est souvent le moyen de déceler les erreurs de couverture. Cependant, ces erreurs sont faciles à corriger (les fournisseurs de données sont ajoutés à la base de sondage ou en sont retirés).

Le degré de sous-dénombrement et de surdénombrement dans une population de référence est déterminé à partir du calcul du taux de sous-dénombrement et du taux de surdénombrement.

Par exemple, si des données sur un ensemble défini d'unités étaient recueillies à l'aide d'une ou de plusieurs bases de sondage, le **taux de sous-dénombrement** (exprimé en pourcentage) serait calculé ainsi :

$$\frac{\text{Unités absentes de la base de sondage mais présentes dans la population de référence}}{\text{Unités présentes dans la population de référence}} \times 100$$

Le **taux de surdénombrement** (exprimé en pourcentage) serait calculé ainsi :

$$\frac{\text{Unités présentes dans la base de sondage mais absentes de la population de référence}}{\text{Unités présentes dans la population de référence}} \times 100$$

L'estimation du nombre d'unités dans la population de référence doit être ajustée en fonction du sous-dénombrement et du surdénombrement, comme l'indique l'exemple ci-dessous.

#### Exemple : calcul de l'erreur de couverture

Nombre d'unités dans la base de sondage	=	1 100
Nombre d'unités manquantes	=	25
Nombre d'unités incluses à tort	=	5

La population de référence est déterminée en fonction des cas de sous-dénombrement et de surdénombrement dans la base de sondage disponible. Ainsi :

$$\begin{aligned} \text{Population de référence} &= (\text{nombre dans la base de sondage}) + (\text{nombre manquant}) - (\text{nombre} \\ &\quad \text{inclus à tort}) \\ &= 1\ 100 + 25 - 5 = 1\ 120 \end{aligned}$$

$$\text{Taux de sous-dénombrement} = \frac{\text{nombre manquant}}{\text{population de référence}} \times 100 = \frac{25}{1\ 120} \times 100 = 2,2 \%$$

$$\text{Taux de surdénombrement} = \frac{\text{nombre inclus à tort}}{\text{population de référence}} \times 100 = \frac{5}{1\ 120} \times 100 = 0,4 \%$$

#### Couverture dans le cas des soumissions volontaires de données

**NOUVEAU**

Les paragraphes précédents traitaient de la couverture lorsque la soumission est obligatoire. Lorsque la soumission est volontaire, la différence entre la **population cible** et la **population de référence** est souvent considérable. Il faut éviter de tirer des conclusions sur la population cible lorsque des données sont manquantes. Les énoncés doivent porter sur la population de référence uniquement, soit les fournisseurs qui soumettent de façon volontaire. Par exemple, si la population cible correspond à l'ensemble de l'équipement médical d'une province et que seules les données concernant les appareils les plus vieux sont soumises sur une base volontaire, les énoncés ne peuvent pas porter sur l'ensemble de la population cible puisque la banque de données n'a pas reçu de données sur les appareils neufs.

Dans les situations de soumission volontaire de données, la population au sujet de laquelle on souhaite obtenir des données doit être clairement définie. Une fois que la base de sondage est établie pour les fournisseurs ayant accepté de fournir des données et qu'un profil contenant les renseignements essentiels est créé à leur intention, il est possible de définir la population de référence et, ainsi, de calculer les taux de surdénombrement et de sous-dénombrement. Il est important d'appliquer régulièrement les procédures de mise à jour de la base de sondage afin de déterminer si les unités font toujours partie du champ d'observation de la population de référence. En outre, les modifications apportées à la base de sondage au fil du temps, soit les ajouts et les suppressions, doivent être consignées dans des documents comme le rapport d'évaluation de la qualité des données (voir la section 4.1) ou la documentation sur la qualité des données à l'intention des utilisateurs (voir la section 4.2).

**Critère 1a** *La population de référence est explicitement mentionnée dans toutes les diffusions.*

Comme il est très important que les utilisateurs des données sachent qui ou qu'est-ce qui fait l'objet d'un examen, la population de référence doit être explicitement mentionnée dans toutes les diffusions. **La diffusion doit informer les utilisateurs de toute différence entre la population de référence et la population cible.** On appelle diffusion tout rapport, toute publication de données ou tout produit découlant d'une banque de données. La diffusion devrait explicitement indiquer si la population de référence a été créée à partir de données soumises de façon obligatoire ou volontaire. Dans le cas des soumissions volontaires, toute entente importante devrait également être clairement précisée. Les questions ou clarifications mineures doivent à tout le moins être mentionnées sous forme de notes en bas de page.

La population de référence pour une diffusion de données peut être différente de celle d'une base de données si la diffusion porte uniquement sur un sous-ensemble de la base de données (p. ex. les données relatives à l'Ontario seulement). S'il peut être nécessaire d'indiquer la population de référence pour la diffusion de données, la population de référence pour la base de données devrait également l'être.

Ce critère est respecté uniquement si la population de référence est mentionnée dans *toutes* les diffusions qui ont eu lieu depuis la dernière évaluation de la qualité des données ou depuis le début des diffusions de données.

**Critère 1b** *Des efforts sont faits pour réduire l'écart entre la population de référence et la population cible.*

Si une différence entre la population de référence et la population cible a été soulevée, il faut expliquer quels efforts ont été faits par la section pour réduire l'écart entre ces deux populations.

Ce critère est respecté si la section a mené des activités visant à réduire l'écart dans la mesure du possible, ou qu'elle est en train de le faire.

**NOUVEAU**

**Critère 2** *Les sources connues de sous-dénombrement et de surdénombrement ont été documentées.*

Les sources connues de sous-dénombrement ou de surdénombrement doivent être documentées sur une base régulière. Tel que le décrit le tableau B, chaque niveau d'observation (unité de la base de sondage) doit être mentionné. S'il est possible de corriger le sous-dénombrement ou le surdénombrement, les corrections aux données doivent être apportées avant la publication des résultats. Il faut également documenter la source de l'erreur pour usage interne. Le surdénombrement peut être corrigé grâce à la suppression des unités hors du champ d'observation ou en double, tandis que le sous-dénombrement peut être corrigé à l'aide d'une source indépendante qui permet de déterminer les unités manquantes ou nouvelles. S'il est impossible de corriger les données, les sources connues de sous-dénombrement ou de surdénombrement devraient être mentionnées dans tous les documents sur la qualité des données.

Ce critère est respecté lorsque les sources connues de l'erreur de couverture sont documentées à l'interne et à l'externe, au besoin.

**Critère 3** *La base de sondage a été validée au moyen d'une comparaison des sources externes et indépendantes.*

Afin de déceler des erreurs dans la base de sondage, il faut parfois la comparer à des sources externes indépendantes de la banque de données. Même si la base de sondage doit être mise à jour périodiquement, la comparaison du contenu de la base de sondage à une source externe indépendante de la documentation de la base de données permet de garantir l'actualité de la base de sondage (p. ex. pour vérifier qu'elle contient les bons établissements de soins de longue durée, il faudrait faire la comparaison avec chaque ministère provincial de la Santé). Dans toute comparaison avec des données externes, les sources externes devraient être crédibles et, dans la mesure du possible, nombreuses.

Généralement, les sources de données sont dites indépendantes lorsqu'elles proviennent d'endroits différents n'ayant aucun lien avec la base de données en question. Dans bon nombre de cas, il est impossible de trouver une source externe d'information complètement indépendante. Par exemple, de nombreuses sources de données ont la même origine, comme les ministères provinciaux de la Santé ou les facultés de médecine, de soins infirmiers ou de pharmacie. S'il est impossible de trouver une source de données indépendante, la base de sondage devrait à tout le moins être vérifiée par les sources ci-dessus.

Dans les cas où aucune source externe au niveau de détail requis n'est trouvée, il est possible d'effectuer une comparaison au niveau agrégé afin de déceler les erreurs dans la base de sondage. Par exemple, si la population de référence correspond aux établissements du Nouveau-Brunswick dotés d'appareils d'IRM et que le nombre ne peut être obtenu qu'auprès d'une source indépendante plutôt qu'à partir d'une liste des établissements, il est possible de comparer le nombre d'établissements de la base de sondage avec celui provenant de la source externe. La concordance des chiffres ne garantit pas l'absence d'erreurs (car les établissements figurant dans la base de sondage ne sont peut-être pas les bons). En revanche, la non-concordance des chiffres signifie qu'une des sources d'information pose problème, lequel devra être réglé avant d'aller plus loin.

Ce critère est respecté si une comparaison annuelle de la base de sondage à des sources externes est effectuée en vue de déceler la présence d'erreurs.

**Critère 4** *Le taux de sous-dénombrement ou de surdénombrement se classe parmi l'une des catégories prédéfinies.*

Ce critère permet d'appliquer une certaine mesure ou évaluation qualitative de l'effet du sous-dénombrement ou du surdénombrement. Voici les lignes directrices à prendre en considération :

Cote possible	Taux de sous-dénombrement ou de surdénombrement (%)
Aucun ou minimal	Moins de 1 %
Modéré	De 1 à 5 %
Significatif	Plus de 5 %
Inconnu	N'a pu être déterminé

Au moment d'attribuer une cote à une base de données, il est nécessaire d'examiner le taux de sous-dénombrement et le taux de surdénombrement séparément et de retenir la cote la plus faible. Par exemple, une base de données avec un surdénombrement minimal et un sous-dénombrement significatif devrait recevoir la cote *significatif*. Il faut également tenir compte des différents niveaux d'observation dans le calcul du sous-dénombrement et du surdénombrement. Chaque unité de la base de sondage doit être évaluée individuellement, comme l'indique le tableau B.

**Ces cotes doivent être utilisées à titre indicatif seulement, car l'effet du sous-dénombrement ou du surdénombrement dépend en grande partie de la quantité et de la répartition des données manquantes.** Par exemple, l'omission d'un hôpital de l'Ontario n'entraînera sans doute pas d'effet significatif sur les estimations provinciales. En revanche, l'omission d'un hôpital de l'Île-du-Prince-Édouard risque d'influer fortement sur l'estimation provinciale. Il faut également savoir que, dans le cas de données sur des patients, un très grand hôpital peut avoir le même effet sur le sous-dénombrement ou le surdénombrement que plusieurs petits hôpitaux.

**Il est très important de noter que le surdénombrement n'annule pas le sous-dénombrement.** Ainsi, une base de données dont les taux de sous-dénombrement et de surdénombrement sont de 5 % présente tout de même une erreur de couverture. Les unités hors du champ d'observation ou en double ne compensent pas les unités manquantes de la base de sondage.

## Saisie et collecte

### Critères

- 5a) *Les pratiques de l'ICIS qui minimisent le fardeau de réponse sont documentées.*
- 5b) *L'ICIS possède de la documentation sur les pratiques des fournisseurs de données qui minimisent le fardeau de réponse.*
- 6 *Il existe des pratiques qui favorisent la collaboration pour la soumission de données.*
- 7 *Il existe des pratiques qui appuient les fournisseurs de données.*
- 8 *Il existe des processus de soumission standard et ils sont suivis par les fournisseurs de données.*
- 9 *Il existe des mesures de contrôle de la qualité de la saisie des données et elles sont mises en œuvre par les fournisseurs de données.*

La caractéristique « saisie et collecte » désigne les pratiques utilisées auprès des fournisseurs de données et durant la saisie des données. On appelle **fournisseur de données** la personne ou l'organisme qui fournit des données à une banque de données. Les fournisseurs de données se chargent habituellement de la **saisie des données**, soit l'entrée à proprement parler des données dans un format utilisable. La **collecte des données** consiste en un rassemblement, à l'ICIS, des données soumises par différents fournisseurs de données dans une banque de données commune.

Prenons l'exemple d'une banque de données sur la santé mentale. Le fournisseur de données peut être défini comme l'établissement de soins de santé mentale ou le ministère de la Santé. La saisie des données correspond à l'abrégé des dossiers de l'établissement réalisé par l'établissement de soins de santé mentale même ou au transfert de ces dossiers au ministère de la Santé par l'établissement. La collecte correspond à la réception du fichier électronique transmis à l'ICIS par l'établissement ou le ministère de la Santé.

La technologie actuelle permet la saisie automatique de l'information à l'aide d'un logiciel créé à l'interne ou par un fournisseur externe. Voici les avantages d'un logiciel de saisie optimal :

- seule l'information relative aux éléments de données pertinents est recueillie (par exemple, il n'est pas nécessaire de recueillir l'âge si la date de naissance est recueillie);
- la saisie des éléments de données obligatoires est imposée;
- les valeurs par défaut ne sont pas définies. Il est nécessaire de distinguer une valeur qui n'est pas présente mais qui devrait l'être, donc une valeur manquante ou inconnue (non-réponse), d'une valeur non applicable;
- des menus déroulants permettent de minimiser les erreurs d'entrée;
- des vérifications intégrées génèrent des messages d'erreur pertinents permettant de repérer les erreurs et d'apporter des modifications rapidement;
- il peut être facilement adapté aux nouvelles exigences, au besoin;
- dans le cas d'un logiciel provenant d'un fournisseur externe, toutes les définitions et les valeurs valides sont fournies dans le manuel d'utilisation.

Une bonne relation avec les fournisseurs de données est primordiale, car elle favorise non seulement la probabilité et l'actualité de la réponse, mais aussi la qualité des données. La question à se poser est la suivante : « Que pouvons-nous faire pour faciliter le plus possible le travail des fournisseurs de données tout en obtenant l'information dont nous avons besoin? »

**Critère 5a**      *Les pratiques de l'ICIS qui minimisent le fardeau de réponse sont documentées.*

Ce critère évalue si des mesures sont appliquées afin que le travail requis par le fournisseur de données soit réduit au minimum sans que la qualité des données en soit compromise. C'est ce qu'on appelle aussi minimiser le **fardeau de réponse**. Plusieurs méthodes peuvent être utilisées, notamment : saisie et soumission électroniques (à l'aide d'un logiciel de saisie, par exemple), calendriers de soumission raisonnables et exclusion d'éléments de données non nécessaires. Toutefois, il est à noter que le fournisseur peut parfois recueillir des données pour ses propres besoins. Dans ce cas, la section de l'ICIS pourrait être dans l'impossibilité d'obtenir toute l'information requise à moins de déposer des demandes supplémentaires auprès du fournisseur.

Ce critère est respecté s'il existe à l'ICIS de la documentation sur les pratiques visant à minimiser le fardeau de réponse.

NOUVEAU

**Critère 5b** *L'ICIS possède de la documentation sur les pratiques des fournisseurs de données qui minimisent le fardeau de réponse.*

Ce critère évalue si des mesures sont appliquées par un organisme (dans le cas où l'ICIS ne saisit ou ne collecte pas directement les données auprès de la source initiale) pour saisir les données, de façon à réduire au minimum le travail réalisé par la source initiale. C'est ce qu'on appelle aussi minimiser le **fardeau de réponse**. Plusieurs méthodes peuvent être utilisées, notamment : saisie et soumission électroniques (à l'aide d'un logiciel de saisie, par exemple), calendriers de soumission raisonnables et exclusion d'éléments de données non nécessaires.

Dans les cas où l'ICIS ne saisit ou ne collecte pas directement les données auprès de la source initiale, ce critère est respecté si une telle documentation est accessible par l'ICIS. Si ce n'est pas le cas, ou si la documentation n'est pas disponible auprès du fournisseur de données, ce critère n'est alors pas respecté.

**Critère 6** *Il existe des pratiques qui favorisent la collaboration pour la soumission de données.*

Les pratiques qui favorisent la collaboration sont importantes, peu importe si la soumission des données est volontaire ou obligatoire. Toutefois, lorsque les fournisseurs de données ont le choix de ne pas répondre, un effort concerté est nécessaire pour obtenir des données de qualité.

Les pratiques qui incitent les fournisseurs à soumettre des données peuvent se résumer à souligner l'importance de la participation, à assurer la confidentialité, à offrir une compensation pour la collaboration (p. ex. publications, formation et rapports spécialisés gratuits) ou à remercier le fournisseur de données par lettre ou par courriel. Ces deux dernières pratiques favoriseront en outre la collaboration future.

Ce critère est respecté s'il existe des pratiques qui favorisent la collaboration.

**Critère 7** *Il existe des pratiques qui appuient les fournisseurs de données.*

Il est essentiel d'offrir un soutien aux fournisseurs de données pour favoriser la soumission rapide et correcte des données. Ce soutien peut être apporté avant et pendant la saisie des données. Avant la saisie des données, le soutien peut prendre la forme de séances d'information et de formation. Il peut aussi s'agir de personnes prêtes à répondre rapidement aux courriels et aux appels téléphoniques des fournisseurs de données. Pendant la saisie des données, les courriels et les appels devraient être traités tout aussi promptement. De plus, un soutien technique et une aide à la codification devraient être offerts aux fournisseurs de données au moyen de documents pertinents, de lignes directrices de codification et d'un manuel de saisie des données.

Ce critère est respecté si des mesures de soutien aux fournisseurs de données sont mises en œuvre avant et pendant la saisie des données.

**Critère 8** *Il existe des processus de soumission standard et ils sont suivis par les fournisseurs de données.*

Les processus de soumission standard facilitent le processus de collecte tant pour le fournisseur de données que pour le personnel de l'ICIS. Ils peuvent également favoriser l'actualité des données et réduire le nombre d'erreurs.

Les processus de soumission standard assurent une collecte de données aussi uniforme que possible entre les fournisseurs. Une banque de données qui reçoit de certains fournisseurs des données soumises annuellement sur papier et, d'autres fournisseurs, des données soumises mensuellement sous forme électronique ne s'appuie pas sur des processus standard. Dans le même ordre d'idées, les éléments de données recueillis devraient être les mêmes dans chaque province et territoire et les éléments de données obligatoires et facultatifs devraient présenter des similarités. Cela facilite l'établissement de comparaisons significatives entre les provinces et les territoires.

Ce critère est respecté si les processus de soumission servant à la collecte des données sont normalisés et suivis par les fournisseurs de données.

**Critère 9** *Il existe des mesures de contrôle de la qualité de la saisie des données et elles sont mises en œuvre par les fournisseurs de données.*

Les mesures de **contrôle de la qualité de la saisie des données** sont exécutées au moment de l'entrée des données. Ces mesures garantissent l'enregistrement correct des données. Il peut s'agir de vérifications lors de la saisie des données, d'une vérification visuelle des données, d'une double saisie ou d'autres procédures.

Les **vérifications** de la saisie des données peuvent nettement améliorer la qualité des données envoyées à l'ICIS, car elles permettent de contrôler ou de corriger les données originales. Les vérifications de la saisie de données consistent généralement à évaluer la validité des éléments de données (p. ex. vérifier si le sexe indiqué est Homme, Femme, Autre ou Inconnu, ou si le numéro d'identification du patient est de la longueur appropriée). De plus, les vérifications de la cohérence examinent le lien entre les éléments de données (p. ex. vérifier, au moyen d'un système de saisie, une intervention qui peut être réalisée uniquement sur des femmes). Des vérifications doivent également avoir lieu au moment de la saisie initiale des données (voir les critères 18 à 21).

Il y a une **vérification visuelle** lorsqu'une deuxième personne examine la concordance entre les données originales et les données saisies, et **double saisie** lorsque deux personnes enregistrent chacune les données pour déceler les écarts. Comme ces procédures exigent souvent beaucoup de ressources, elles sont généralement réalisées à partir d'échantillons. Il est à noter que le personnel de la Division de la qualité des données peut aider les sections de l'ICIS à établir les paramètres de ces vérifications.

Ce critère est respecté si des mesures de contrôle de la qualité sont utilisées lors de la saisie des données.

## Non-réponse totale

### Critères

- 10 *L'ampleur de la non-réponse totale est mentionnée dans la documentation sur la qualité des données.*
- 11 *Le nombre d'enregistrements reçus pour les unités déclarantes est examiné afin que les valeurs inhabituelles puissent être décelées.*
- 12 *L'ampleur de la non-réponse totale se classe parmi l'une des catégories prédéterminées.*

Il y a **non-réponse totale** lorsque les réponses d'unités entières (provinces, régions sanitaires, établissements, dossiers de patients, etc.) sont absentes de la banque de données. On confond souvent non-réponse totale et sous-dénombrement, car les deux se produisent lorsque des unités complètes sont manquantes. Par exemple, si une des unités de la base de sondage d'une banque de données ne soumet pas de données, il s'agit d'un cas de non-réponse. Par contre, si cette unité ne fait pas partie de la base de sondage mais qu'elle fait partie de la population de référence, il s'agit d'un cas de sous-dénombrement. Au moment d'examiner les unités de la base de sondage, il faut passer en revue les différents niveaux d'observation du tableau B afin de déterminer les unités qui seront déclarées à la présente étape.

De par la nature des bases de données administratives, la collecte de données se fait souvent par l'entremise des fournisseurs de données, ce qui peut donner lieu à différents niveaux de non-réponse (patient, dispensateur de services, établissement, région sanitaire ou province). Par exemple, si on s'attend à ce que les ministères de la Santé soumettent des données concernant l'ensemble des demandes de remboursement de médicaments faites auprès de leurs programmes publics de médicaments (population de référence), il se peut que les données relatives à certains programmes ne soit pas soumises, ou encore que seule une partie des données sur les demandes de remboursement d'un programme quelconque ne le soit. Dans les deux cas, il s'agit d'exemples de non-réponse totale. Il est important d'évaluer la quantité de données non soumises par un fournisseur. La non-réponse totale devrait être limitée le plus possible afin de :

- réduire la perte de renseignements précieux;
- prévenir les résultats biaisés lorsque les unités pour lesquelles l'ICIS n'a pas reçu de données diffèrent de celles pour lesquelles l'ICIS en a reçues.

Le personnel de l'ICIS calcule habituellement le **taux de non-réponse totale** plutôt que son complément, le taux de réponse totale. Le calcul du taux de réponse totale peut être fondé sur le nombre d'**unités de la base de sondage** ou d'**unités d'analyse**. L'unité de la base de sondage est l'unité sur laquelle se fonde la base de sondage d'une banque de données (p. ex. patient, dispensateur de services, établissement, région sanitaire, province ou autre fournisseur de données). L'unité d'analyse, quant à elle, peut correspondre à l'unité de la base de sondage ou à une unité faisant partie de celle-ci (p. ex. les patients au sein d'un établissement). Consultez le tableau B pour connaître d'autres exemples de ces niveaux d'observation. Le taux de non-réponse des **unités de la base de sondage** (exprimé en pourcentage) se calcule ainsi :

$$\frac{\text{Nombre d'unités de la base de sondage n'ayant pas soumis de données} \times 100}{\text{Nombre d'unités dans la base de sondage}}$$

Le taux de non-réponse des **unités d'analyse** par rapport à l'ensemble des unités de la base de sondage (exprimé en pourcentage) se calcule ainsi :

$$\frac{\text{Nombre d'unités n'ayant pas soumis de données pour chaque unité de la base de sondage} \times 100}{\text{Nombre d'unités qui auraient dû soumettre des données pour chaque unité de la base de sondage}}$$

Le taux de réponse totale (exprimé en pourcentage) correspond tout simplement à 100 % - le taux de non-réponse totale.

Selon la banque de données, il peut être approprié de calculer le taux de non-réponse de sous-groupes, comme la taille, le type, la région ou la province, en plus du taux de non-réponse global. Les données manquantes d'une unité de grande envergure de la base de sondage peuvent entraîner la perte d'unités d'analyse beaucoup plus nombreuses que les données manquantes d'une unité de petite envergure de la base de sondage. Le taux approximatif de non-réponse des unités peut être calculé grâce à la comparaison du nombre global d'unités d'analyse manquantes au nombre d'unités d'analyse attendues. Ce calcul peut devenir assez complexe. Il est donc recommandé de demander l'aide du personnel de la Division de la qualité des données l'ICIS.

Bien que la non-réponse des établissements puisse être facile à détecter, il n'est pas toujours possible de déterminer si les unités de la base de sondage ont fourni toutes les unités d'analyse requises. Toutefois, il est assez facile de relever les variations importantes dans le nombre d'unités soumis par une unité de la base de sondage. Ces variations peuvent indiquer un fort taux de non-réponse ou l'inclusion d'unités incorrectes et devraient donc être examinées. Afin de repérer les variations inhabituelles, il est nécessaire d'effectuer une surveillance régulière du nombre d'unités d'analyse qui n'ont pas été soumises par chaque unité de la base de sondage.

À partir de l'établissement de comparaisons avec les années précédentes ou de la surveillance effectuée, la non-réponse totale est résolue efficacement au moyen de mesures de suivi. Par exemple, le suivi d'une unité de la base de sondage qui ne soumet pas de données est nécessaire si l'on veut s'assurer qu'elle fait toujours partie du champ d'observation. Un suivi devrait également être effectué auprès des unités déclarantes lorsqu'une comparaison par rapport à une période précédente révèle un nombre inhabituel d'unités d'analyse.

Une bonne communication (séances d'information et de formation, manuel de saisie des données, etc.) avec les fournisseurs de données avant le début de la période de collecte de données peut réduire le taux de non-réponse. En outre, il est possible d'alléger le fardeau de réponse des fournisseurs de données en leur demandant de préciser ce qui leur faciliterait la tâche et en mettant en œuvre les mesures nécessaires. Il importe aussi de tirer des leçons des erreurs commises et de les documenter. Le fait d'allouer suffisamment de temps aux fournisseurs pour soumettre les données et entreprendre les activités de suivi auprès d'eux contribue également à diminuer le taux de non-réponse.

**Critère 10** *L'ampleur de la non-réponse totale est mentionnée dans la documentation sur la qualité des données.*

Il est important de préciser l'ampleur de la non-réponse totale des unités de la base de sondage dans la documentation sur la qualité des données afin que les utilisateurs puissent évaluer l'intégralité des données. Comme il a été précisé ci-dessus, l'unité de la base de sondage peut se référer aux patients, aux dispensateurs de services, aux régions sanitaires, aux provinces ou à tout autre fournisseur de données. Si les taux de non-réponse varient fortement selon la province ou la région, ils doivent être publiés selon ces niveaux.

Ce critère est respecté si l'ampleur de la non-réponse totale des unités de la base de sondage est précisée dans la documentation sur la qualité des données à un niveau de détail pertinent pour la plupart des analyses.

**Critère 11** *Le nombre d'enregistrements reçus pour les unités déclarantes est examiné afin que les valeurs inhabituelles puissent être décelées.*

Pour déceler la non-réponse, il est important de faire le suivi du nombre d'unités déclarantes au fil du temps afin de relever les valeurs inhabituelles. Par exemple, si un fournisseur de données soumet chaque mois des enregistrements relativement à environ 1 000 unités et que ce nombre passe soudainement à 1 500 ou à 500 unités, un examen rigoureux s'impose. Une variation du nombre d'unités soumises n'indique pas nécessairement qu'il y a un problème, puisque bien des raisons peuvent l'expliquer (p. ex. soumission tardive des enregistrements, agrandissement ou fermeture temporaire de l'établissement, nouvelles pratiques d'embauche, épidémie causant une réduction des visites à l'hôpital, etc.).

Ce critère est respecté si un suivi au fil du temps du nombre d'enregistrements soumis par les unités déclarantes de la base de sondage est réalisé en vue de déceler les valeurs inhabituelles.

**Critère 12** *L'ampleur de la non-réponse totale se classe parmi l'une des catégories prédéterminées.*

Tel que mentionné précédemment, la non-réponse peut se produire à plusieurs niveaux. Comme ce critère tient autant compte du niveau de la base de sondage que du niveau de l'unité d'analyse, il est recommandé que la non-réponse soit calculée pour chacun d'eux. À l'ICIS, c'est habituellement le taux de non-réponse totale qui est calculé plutôt que son complément, le taux de réponse totale.

Le taux de non-réponse totale au niveau de la base de sondage (exprimé en pourcentage) se calcule ainsi :

$$\frac{\text{Nombre d'unités de la base de sondage n'ayant pas soumis de données} \times 100}{\text{Nombre d'unités dans la base de sondage}}$$

Le taux de non-réponse totale au niveau de l'unité d'analyse (exprimé en pourcentage) se calcule ainsi :

$$\frac{\text{Nombre d'unités d'analyse n'ayant pas soumis de données} \times 100}{\text{Nombre total d'unités d'analyse attendues ou connues}}$$

Le tableau suivant présente les cotes d'évaluation proposées pour le taux de non-réponse au niveau de l'unité de la base de sondage ou de l'unité d'analyse. Si la non-réponse se situe bien au-delà du taux souhaité parce que les unités de la base de sondage n'ayant pas répondu sont de grande envergure, les cotes peuvent être modifiées de façon à ce qu'elles reflètent plus fidèlement la gravité des problèmes créés par les données manquantes. Idéalement, le taux de non-réponse totale devrait être déterminé par les personnes qui connaissent le mieux les unités de la base de sondage (c.-à-d. le fournisseur de données ou le personnel de l'ICIS). Si le personnel de l'ICIS ne connaît pas l'ensemble des unités de la base de sondage, il est important que l'ICIS rassemble les renseignements pertinents (p. ex. le statut du répondant) sur chaque unité auprès du fournisseur de données afin de déterminer avec exactitude le taux de non-réponse totale.

<b>Cote proposée</b>	<b>Taux de non-réponse totale (%)</b>
Aucun ou minimal	Moins de 2 %
Modéré	De 2 à 10 %
Significatif	Plus de 10 %

## Non-réponse partielle (ou non-réponse par question)

### Critères

13 *La non-réponse partielle est connue.*

14 *L'ampleur de la non-réponse partielle se classe parmi l'une des catégories prédéterminées.*

Il y a **non-réponse partielle** (également appelée **non-réponse par question**) lorsqu'une unité (unité de la base de sondage ou unité d'analyse) pour laquelle des données ont été reçues ne contient que des renseignements partiels au sujet des éléments de données ou des transactions (dans le cas des banques de données financières). La non-réponse partielle ne tient pas compte des éléments de données non applicables à l'unité, autrement dit lorsque l'information relative à certains éléments de données est manquante. Le terme *manquant* fait référence aux éléments de données pour lesquels des renseignements étaient attendus, mais dont le champ est demeuré vide dans le fichier de données. Les données sont considérées comme manquantes même si des valeurs par défaut (soit 0 ou 99) sont utilisées pour codifier un élément non déclaré en tant que non-réponse. Aux fins d'établissement de la non-réponse partielle, les éléments de données auxquels la valeur 0 a été attribuée ne devraient pas être considérés comme manquants. La non-réponse partielle diffère de la non-réponse totale en ce que cette dernière porte sur les réponses d'unités entières (comme il en a été question aux critères 10 à 12), tandis que la non-réponse partielle porte sur les éléments de données faisant partie des unités.

Le taux de non-réponse partielle est assez simple à calculer lorsqu'il est possible de déterminer si les éléments de données sont manquants ou s'ils sont non applicables. Si l'on demandait à des hôpitaux d'indiquer leur nombre de lits en psychiatrie, un champ vide pourrait signifier que l'établissement ne dispose pas de lits en psychiatrie (non applicable) ou qu'il n'a tout simplement pas répondu (manquant). Durant le processus de vérification, il est recommandé d'utiliser un élément de données supplémentaire pour différencier les valeurs manquantes de celles qui ne s'appliquent pas. Par exemple, selon la norme acceptée dans le Dictionnaire de données de l'ICIS, la valeur 7 (non recueilli) ou 9 (inconnu) est attribuée aux données manquantes, tandis que la valeur 8 est réservée aux données non applicables. En l'absence d'un élément de données supplémentaire, il importe de distinguer les réponses manquantes des réponses non applicables avant le calcul du taux de non-réponse partielle, puisque les réponses non applicables doivent en être exclues.

Le personnel de l'ICIS calcule habituellement le **taux de non-réponse partielle** plutôt que son complément, le taux de réponse partielle.

Le taux de non-réponse partielle pour un élément de données (exprimé en pourcentage) se calcule ainsi :

$$\frac{\text{Nombre d'unités pour lesquelles des données relatives à un élément de données n'ont pas été fournies}}{\text{Nombre d'unités qui auraient dû fournir les données relatives à l'élément de données}} \times 100$$

Il est à noter que les unités (unités de la base de sondage ou unités d'analyse) contenant un élément de données non applicable doivent être exclues du numérateur et du dénominateur lors du calcul du taux de non-réponse partielle relatif à cet élément de données.

L'exemple qui suit explique plus en détail le concept de non-réponse partielle. Cent unités (des établissements) provenant de la base de sondage d'une banque de données ont été invitées à répondre à deux questions : Combien de pédiatres travaillent dans votre établissement? et Combien d'infirmières travaillent dans votre établissement? Les réponses à ces questions ont indiqué que 70 établissements comptent des pédiatres et des infirmières et que 5 ne comptent que des infirmières (il n'y a jamais de pédiatres). Les autres établissements, pour lesquels nous savons que des pédiatres et des infirmières y sont en fonction, n'ont soumis aucune réponse.

Le taux de non-réponse partielle relatif au nombre de pédiatres est donc de 25/95, soit 26,3 % (les 5 unités ne comptant pas de pédiatres ont été exclues du numérateur et du dénominateur), et le taux de non-réponse partielle relatif au nombre d'infirmières est de 25/100, soit 25 %.

Il faut savoir que le taux de non-réponse partielle est calculé par rapport au nombre d'unités qui déclarent l'élément de données (c'est-à-dire que l'élément de données est applicable). Dans l'exemple ci-dessus, le taux de non-réponse partielle concernant le nombre de pédiatres est de 26,3 %, mais il y a absence de données sur les pédiatres pour 30 % (30/100) des établissements. Le taux de non-réponse partielle ne brosse donc pas un portrait complet de l'intégralité des données, d'où l'importance de tenir compte également de la non-réponse totale. Une base de données affichant un taux de non-réponse partielle de 0 % pour tous les éléments de données pourrait présenter beaucoup de données manquantes si le taux de réponse totale est élevé. Dans le même ordre d'idées, une base de données affichant un faible taux de non-réponse totale et un fort taux de non-réponse partielle aura également beaucoup de données manquantes, particulièrement étant donné que des unités complètes sont rendues inutiles aux fins d'analyse lorsque de nombreux éléments de données sont manquants.

Ce critère est respecté si le taux de non-réponse partielle relatif à tous les éléments de données de base (soit ceux qui sont régulièrement utilisés dans les analyses de données) est calculé.

**Critère 13** *La non-réponse partielle est connue.*

Comme il a été mentionné précédemment, pour déterminer l'étendue de la non-réponse partielle dans une base de données, il est important de pouvoir distinguer un champ vide d'une non-réponse.

Les éléments de données sont habituellement signalés lors du processus de vérification au moyen de la création d'une valeur supplémentaire relative à l'élément de données existant (p. ex. toute valeur qui n'est pas une valeur de réponse possible) ou au moyen de la création d'un nouvel élément de données. Par exemple, selon la norme acceptée dans le Dictionnaire de données de l'ICIS, la valeur 7 (non recueilli) ou 9 (inconnu) est attribuée aux données manquantes, tandis que la valeur 8 est réservée aux données non applicables. Il est ainsi facile de différencier un élément de données manquant d'un élément de données non applicable.

Ce critère est respecté si la non-réponse partielle peut être connue pour tous les **éléments de données de base** et si elle est signalée lorsque l'élément de données est requis dans certaines conditions uniquement. Un élément de données de base est un élément régulièrement utilisé dans les analyses de données.

**Critère 14** *L'ampleur de la non-réponse partielle se classe parmi l'une des catégories prédéterminées.*

Ce critère met l'accent uniquement sur le volume ou l'ampleur des éléments de données présentant une non-réponse partielle. L'incidence de l'ampleur de la non-réponse partielle dépend de nombreux facteurs, notamment l'importance des réponses manquantes aux éléments de données (de base ou non) et la présence d'une tendance dans les valeurs manquantes. Idéalement, les valeurs manquantes devraient être « manquantes de façon totalement aléatoire » pour éviter qu'un biais soit introduit dans les résultats. Les données sont considérées comme manquantes de façon complètement aléatoire si les valeurs sont aléatoirement absentes et qu'elles ne sont liées à aucun autre élément de données.

Le niveau de non-réponse de chaque élément de données de base devrait être évalué. Un élément de données de base est un élément régulièrement utilisé dans les analyses. Au moment d'évaluer le niveau de non-réponse partielle, il faut tenir compte de l'élément de données de base qui a le plus fort taux de non-réponse. Les cotes ci-dessus ne sont que des suggestions. Si certains éléments de données sont particulièrement importants ou qu'ils ne sont pas manquants de façon complètement aléatoire, il est possible de modifier la cote pour qu'elle reflète plus fidèlement la gravité des problèmes créés par les données manquantes.

Les cotes proposées sont les suivantes :

Cote proposée	Taux de non-réponse partielle (%)
Aucun ou minimal	Moins de 2 %
Modéré	De 2 à 5 %
Significatif	Plus de 5 %

## Erreur de mesure, biais et cohérence

### Critères

- 15 *Le niveau de l'erreur de mesure se classe parmi l'une des catégories prédéterminées.*
- 16 *Le niveau du biais n'est pas significatif.*
- 17 *Le degré de problèmes de cohérence se classe parmi l'une des catégories prédéterminées.*

Des erreurs peuvent se produire pour de nombreuses raisons. Par ailleurs, il peut souvent être difficile de les regrouper dans le but d'en évaluer facilement les différentes causes. Afin de simplifier cette tâche, le cadre divise l'évaluation des erreurs en trois composantes qui se chevauchent :

1. *erreur de mesure* : erreur causée lorsque la réponse ou le code attribué à un élément de données est incorrect.
2. *biais* : évalue dans quelle mesure l'écart entre les valeurs déclarées et celles qui auraient dû l'être se produit de façon systématique.
3. *cohérence* : évalue la variation susceptible de se produire si des mesures répétées étaient réalisées.

Les paragraphes suivants expliquent plus en détail ces trois composantes.

Une **erreur de mesure** élevée peut indiquer la présence d'un certain nombre de problèmes à régler. Il peut s'agir de définitions peu claires, d'un manque de formation causant la multiplication des interprétations et la variabilité des réponses aux entrées subjectives, la survérification des données ou une faiblesse dans le processus de collecte des données. La mise en œuvre de processus automatisés mis à l'essai et révisés de même qu'une documentation et une formation adéquates contribuent à la réduction des erreurs de mesure.

Il peut être très difficile de prouver l'*existence* d'un **biais** sans réaliser des études spéciales, comme une étude de seconde saisie (dont il sera question plus loin). Il est toutefois relativement facile de déceler les biais *possibles*. Les biais possibles peuvent être détectés lorsque l'on sait qu'il y a couverture ou réponse incomplète, ou encore erreurs de traitement ou d'échantillonnage. Les erreurs de couverture, qui consiste en des omissions, des inclusions erronées et des doubles dans la base de sondage, peuvent entraîner un biais positif ou négatif et leur effet peut varier selon les sous-groupes de la population cible. Il y a erreur de non-réponse lorsque la banque de données n'obtient pas de réponse à une ou possiblement à toutes les questions, ce qui est susceptible de causer un biais si les caractéristiques d'intérêt des non-répondants diffèrent de celles des répondants. Les erreurs de traitement, qui peuvent se produire lors d'étapes suivant la saisie et la collecte, comme la vérification des données et la mise en tableau, peuvent entraîner la production de résultats biaisés. Dans le cas d'un échantillon aléatoire sélectionné à partir de la population cible, un biais peut être introduit dans les estimations si le sondage est fondé sur un échantillon de la population plutôt que sur l'ensemble de la population, ou encore si l'échantillon aléatoire n'est pas adéquat. De plus, dans les enquêtes par sondage, les erreurs d'estimation peuvent introduire un biais selon la méthode d'estimation choisie.

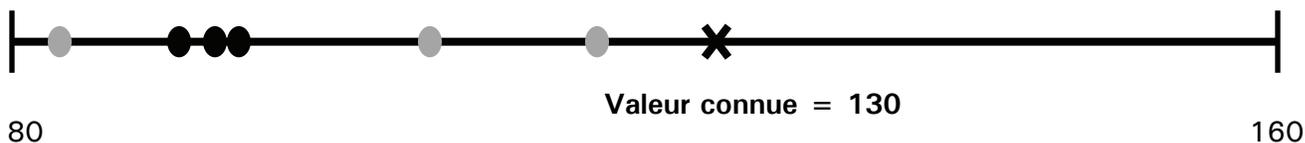
Au moment d'examiner les biais, il est important de tenir compte du **biais corrélé**, qui se produit lorsqu'il existe une corrélation entre deux éléments de données. Par exemple, un biais corrélé peut être introduit lorsque la durée d'observation des patients est corrélée aux résultats que ces derniers ont obtenus. Dans une telle situation, il est possible que l'imposition d'une période d'admissibilité minimale se solde par l'omission de patients dont l'admissibilité est de courte durée. Si la raison du non-respect du critère d'admissibilité minimale est corrélée aux résultats, par exemple un décès, la conclusion peut être biaisée. De façon similaire, si la durée de la période d'observation varie d'un patient à l'autre, il se peut qu'un biais soit introduit en raison de l'impossibilité d'observer un résultat une fois la période d'observation terminée. Même si les biais corrélés sont parfois plus complexes que les biais non corrélés, ils sont souvent plus faciles à déceler parce qu'il est possible de comparer les valeurs d'un élément de données à l'autre et ainsi de relever les écarts.

La **cohérence** mesure la variation dans les réponses lors de mesures répétées, ce qu'on appelle parfois aussi la fiabilité. Les éléments de données subjectifs (comme le niveau d'invalidité sur une échelle de 1 à 5 ou le type de diagnostic) sont ceux dont les réponses peuvent être erronées. La cohérence ne s'applique pas uniquement aux éléments de données subjectifs; elle peut aussi être un facteur dans les éléments de données comportant une erreur de mesure (p. ex. une mesure de la taille incorrecte). La cohérence évalue dans quelle mesure la subjectivité des codificateurs influe sur la codification. Plusieurs techniques statistiques permettent de mesurer la cohérence de la codification, comme le taux de concordance ou la statistique kappa. La statistique (ou le coefficient) kappa mesure la concordance par paire auprès d'un ensemble de codificateurs qui font des jugements et corrige la concordance aléatoire prévue. Le coefficient se calcule ainsi :

$$k = \frac{(\text{nombre de concordances observées} - \text{nombre de concordances prévues})}{(\text{nombre total de paires de répondants} - \text{nombre de concordances prévues})}$$

Les estimations résultantes et les tests de concordance réalisés auprès de multiples codificateurs sont appropriés lorsque les réponses se situent sur une échelle nominale ou ordinale. Veuillez consulter le personnel de la Division de la qualité des données pour savoir quelle technique convient le mieux aux données.

L'exemple suivant illustre plus en détail la différence entre biais et cohérence. Il montre les résultats de la prise de la pression artérielle d'un patient par deux infirmières au cours d'une période déterminée, chaque infirmière ayant pris la pression trois fois.



Comme l'indique l'illustration, les lectures réalisées par l'infirmière représentée par les cercles gris pâle sont **moins cohérentes** que celles effectuées par l'infirmière représentée par les cercles noirs, puisque les mesures de la première sont plus dispersées. L'infirmière représentée par les cercles noirs est donc **plus cohérente** que l'autre. En revanche, l'infirmière représentée par les cercles gris pâle a un résultat **moins biaisé** puisqu'en moyenne, les lectures qu'elle a effectuées se rapprochent davantage de la valeur connue que celles de l'autre infirmière.

Même si le niveau exact d'erreur de mesure, de biais et de cohérence des éléments de données est souvent inconnu, des études spéciales telles que les études de secondes saisies devraient être envisagées. Dans une étude de seconde saisie, un échantillon d'enregistrements (p. ex. des dossiers de patients) est sélectionné, puis recodifié. Les données codifiées à l'origine sont comparées aux données recueillies dans le cadre de l'étude. Les divergences entre les valeurs ainsi que les raisons de ces divergences sont indiquées. L'erreur de mesure est évaluée au moyen du taux de divergence entre la base de données et l'analyste de seconde saisie. Le biais est quant à lui évalué par un examen des taux de divergence en tant que tout : si les erreurs sont systématiques, les résultats seront trompeurs et pourraient devoir être ajustés. La cohérence est évaluée au moyen de mesures répétées, comme les vérifications inter-évaluateur ou intra-évaluateur. Une vérification inter-évaluateur consiste en la codification d'un même dossier par deux analystes de seconde saisie. Cette vérification permet de déterminer si les réponses sont cohérentes entre ceux-ci. Dans une vérification intra-évaluateur, un seul analyste codifie un même dossier deux fois. Cette vérification permet de connaître la capacité de cette personne à enregistrer de l'information de façon cohérente.

En raison de la limitation des ressources, il n'est pas toujours possible de réaliser des études spéciales comme des études de seconde saisie dans le cadre d'une publication de données. Toutefois, le personnel affecté à la banque de données a souvent une idée de la qualité des éléments de données qu'elle contient et est en mesure de réaliser une première évaluation de ces trois types d'erreur.

**Critère 15** *Le niveau de l'erreur de mesure se classe parmi l'une des catégories prédéterminées.*

Le plus souvent, la quantité d'erreurs dans les éléments de données d'une base de données est évaluée grâce à des études de seconde saisie ou à d'autres études spéciales (et habituellement rétrospectives), mais il est aussi possible de le faire lors du processus d'élaboration de la base de données. Le niveau d'erreur est souvent exprimé comme un taux d'erreur ou de divergence, mesuré par le pourcentage de cas codifiés incorrectement à chaque élément de données. Comme cette approche suppose qu'une bonne réponse est possible, les taux d'erreur ne s'appliquent habituellement pas aux éléments de données subjectifs.

Si une étude de seconde saisie a été effectuée, il faut examiner les résultats relatifs aux éléments de données de base et les comparer au tableau ci-dessous afin d'obtenir une cote. La cote à utiliser devrait être celle de l'élément de données de base présentant le taux d'erreur le plus élevé. Cependant, si de nombreux éléments de données affichent des taux d'erreur substantiels ou si les éléments de données sont spécialement importants ou non importants, la cote peut être ajustée en conséquence.

Cote proposée	Taux d'erreur relatif aux variables non subjectives (%)
Aucun ou minimal	De 0 à moins de 5 %
Modéré	De 5 à 10 %
Significatif	Plus de 10 %

Si le niveau d'erreur n'est pas estimé au moyen d'une étude de la qualité des données, le critère ne reçoit pas nécessairement la cote *inconnu*. Dans bien des cas, le personnel de la banque de données n'a pas besoin de réaliser d'étude spéciale; il est déjà conscient des problèmes ou a une idée de la quantité d'erreurs dans les données. Cette connaissance doit être appuyée par d'autres renseignements pour qu'un niveau soit attribué; l'attribution ne doit pas être faite en fonction d'un taux d'erreur numérique donné. Si le niveau d'erreur n'a pas été évalué par une étude et que le personnel de la banque de données est incapable de l'évaluer qualitativement, la cote *inconnu* devrait être attribuée.

#### **Critère 16** *Le niveau du biais n'est pas significatif.*

Le critère d'erreur de mesure évalue la quantité d'erreurs dans les éléments de données non subjectifs de la base de données, tandis que le critère de biais est conçu pour déterminer si les écarts dans les valeurs déclarées sont systématiques. Toutefois, le biais peut s'appliquer tant aux éléments de données subjectifs que non subjectifs. Par exemple, la réponse qu'un patient donne à son dentiste qui lui demande combien de fois par semaine il utilise la soie dentaire (un élément de données non subjectif) est souvent une réponse biaisée, parce que les gens ont tendance à exagérer ce nombre pour faire plaisir à leur dentiste. De la même façon, une personne qui essaie de recevoir une indemnité ou de s'attirer la sympathie à la suite d'une blessure peut être portée à exagérer sa douleur, qui est un élément de données subjectif.

Ce critère est difficile à évaluer à cause de la complexité à prouver la présence d'un biais. Par conséquent, l'évaluation se fera d'après la présence réelle ou présumée d'un biais important dans les données. Les biais (ou les biais possibles) dans les données peuvent être décelés grâce à la comparaison des estimations à des sources externes, à des comparaisons internes permettant de détecter les biais corrélés (valeurs par province, hôpital, etc.) et à la vérification des enregistrements au moyen d'une seconde saisie. Si un biais (ou un biais corrélé) réel ou présumé est suffisamment important pour influencer considérablement sur les estimations, le niveau de biais doit être coté comme *non respecté*. En revanche, s'il n'y a aucun biais réel ou présumé, le niveau de biais doit être coté comme *respecté*. Autrement, le critère devrait être coté comme *inconnu*.

Ce critère est respecté si rien ne prouve ou ne laisse croire qu'il existe un biais suffisamment important pour influencer sur les totaux, le nombre d'occurrences ou les estimations (dans le cas d'une population échantillonnée).

**Critère 17** *Le degré de problèmes de cohérence se classe parmi l'une des catégories prédéterminées.*

La cohérence est source de préoccupation pour toutes les banques de données puisque la cohérence de chaque élément de données dépend grandement de l'opinion ou de l'interprétation des codificateurs, des infirmières et des personnes inscrites, par exemple. Il faut tenir compte de la cohérence des mesures de chacune de ces personnes de même que de la cohérence des mesures entre elles. Comme pour le biais et l'erreur de mesure, la cohérence est le plus souvent évaluée au moyen d'études de seconde saisie, mais elle peut également l'être pendant l'élaboration de la base de données. Il est également possible d'effectuer des vérifications au hasard pour mesurer la cohérence pendant la collecte des données. Comme il a été mentionné précédemment, plusieurs techniques statistiques permettent de mesurer la cohérence et l'exactitude de la codification. Veuillez consulter le personnel de la Division de la qualité des données pour connaître la technique la plus appropriée en fonction des données en question et approfondir l'analyse et l'interprétation.

Si une étude spéciale de la qualité des données a été effectuée, il faut examiner les résultats relatifs aux éléments de données subjectifs et les comparer au tableau ci-dessous afin d'obtenir une cote. La cote applicable aux éléments de données présentant le plus faible niveau de cohérence devrait être utilisée. Cependant, si de nombreux éléments de données comportent des problèmes de cohérence ou si les éléments de données sont spécialement importants ou non importants, il est possible d'ajuster la cote en conséquence.

Cote proposée	Taux de divergence (%)	Statistique kappa
Aucun ou minimal	De 0 à moins de 5 %	De 0,81 à 1,00
Modéré	De 5 à 10 %	De 0,50 à 0,80
Significatif	Plus de 10 %	De -1,00 à 0,49

Si le niveau de cohérence n'est pas estimé au moyen d'une étude de la qualité des données, le critère ne reçoit pas nécessairement la cote *inconnu*. Dans bien des cas, le personnel affecté à la base de données est conscient des problèmes ou a une idée de la fiabilité des données. Cette connaissance doit être appuyée par d'autres renseignements pour qu'un niveau soit attribué; l'attribution ne doit pas être faite en fonction d'un taux d'erreur numérique donné. Si le niveau de cohérence n'a pas été évalué au moyen d'une étude et que le personnel de la banque de données est incapable de l'évaluer qualitativement, la cote *inconnu* devrait être attribuée au critère.

## Vérification et imputation

### Critères

- 18 *Des contrôles de la validité sont réalisés pour chacun des éléments de données et toutes les données invalides sont signalées.*
- 19 *Les règles de vérification et l'imputation sont logiques et appliquées de façon uniforme.*
- 20 *Les rapports de vérification destinés aux utilisateurs sont faciles à utiliser et à comprendre.*
- 21 *Le processus d'imputation est automatisé et conforme aux règles de vérification.*

À l'ICIS, la **vérification** correspond à la mise en œuvre de contrôles visant à cerner les unités ou les éléments de données manquants, invalides ou incohérents afin de connaître les éléments possiblement erronés. L'**imputation** consiste à déterminer et à attribuer des valeurs de remplacement afin de résoudre les problèmes de données relevés à l'étape de la vérification (données manquantes, invalides ou incohérentes). La vérification et l'imputation peuvent faire l'objet d'un seul processus ou de deux processus distincts.

La vérification des données est parfois une tâche complexe. La mise en place de vérifications appropriées représente un investissement dans la qualité des données. En effet, un programme de vérification efficace peut déceler de nombreuses erreurs dans les données qui seraient passées inaperçues autrement. Au moment de mettre en place des vérifications, il est important de se demander quel type de vérifications est approprié et d'en déterminer les paramètres. Après la collecte de données, l'une des premières étapes d'un programme de vérification devrait être de relever les doubles. Un double se produit par exemple lorsqu'un établissement transmet de l'information deux fois au cours de la même journée, ou lorsque de l'information mise à jour relative à une unité est soumise mais que l'information précédente n'est pas retirée de la base de données. Comment relever les doubles? Il faut d'abord déterminer les éléments de données qui permettent d'identifier une unité de façon unique. Ensuite, il faut vérifier que les unités ne sont pas déjà présentes dans la base de données au moment de la réalisation de mises à jour. En outre, un registre des rejets à la vérification et des unités de suivi qui n'ont pas été soumises de nouveau devrait être créé.

Une fois que les doubles ont été résolus, la vérification peut se poursuivre. Il existe essentiellement trois types de vérifications à prendre en considération : vérifications de la validité, vérifications de la cohérence et vérifications de la répartition. Les vérifications de la validité permettent de relever les cas où des éléments de données obligatoires présentent des valeurs manquantes, où le format des réponses est incorrect et où les valeurs des réponses ne correspondent pas à l'une des valeurs attendues. Les vérifications de la cohérence détectent les liens improbables entre les éléments de données (p. ex. un homme donnant naissance à un enfant). Enfin, les vérifications de la répartition cernent les valeurs aberrantes dans la répartition des données en examinant l'ensemble de données complet (p. ex. des temps d'attente considérablement plus longs que la moyenne). Des vérifications de la validité et des vérifications de la cohérence simples peuvent être réalisées lors de la saisie des données (si elles sont intégrées au logiciel de saisie) tandis que des vérifications de la cohérence et de la répartition plus complexes peuvent avoir lieu lors du traitement des données. Il est recommandé de consulter le personnel de la Division de la qualité des données avant de réaliser des vérifications complexes et pour vérifier les paramètres utilisés.

Lorsqu'une unité ou un élément de données fait l'objet d'un signalement à la suite de vérifications lors du traitement des données, il importe de faire un suivi auprès du fournisseur de données avant la fin de la période de collecte. Autrement, l'imputation devrait être envisagée. Une fois la période de collecte terminée, l'imputation peut servir à déterminer les valeurs de remplacement nécessaires pour les valeurs manquantes, invalides ou incohérentes.

À la suite de toute imputation, il faut s'assurer que les données non imputées puissent être différenciées des données imputées. Pour ce faire, il est possible entre autres de créer un élément de données supplémentaire (soit un signalement indiquant que l'imputation a été effectuée ou un élément de données qui contient les valeurs imputées). Par ailleurs, il est judicieux de prendre l'habitude de surveiller le nombre d'enregistrements imputés de même que le nombre de fois qu'une méthode d'imputation (imputation rétrospective, imputation moyenne, etc.) est utilisée. Ce suivi peut aider à déterminer si des modifications doivent être apportées pour améliorer la qualité des données. Lorsqu'une imputation est effectuée, il est également recommandé de mettre à nouveau en œuvre les différentes vérifications afin de s'assurer que l'unité ou l'élément de données imputé est exempt d'erreur et cohérent en soi.

L'imputation logique (ou déductive) et l'imputation rétrospective sont deux méthodes couramment utilisées. Dans le cas de l'imputation logique, il n'existe qu'une seule valeur possible pour une donnée manquante ou invalide (p. ex. si la date d'inscription et la date de la fin de la visite sont la même, la seule valeur possible pour la date de triage est également cette date). L'imputation rétrospective consiste quant à elle à reporter pour la période actuelle les données d'une période précédente.

**Critère 18** *Des contrôles de la validité sont réalisés pour chacun des éléments de données et toutes les données invalides sont signalées.*

Les contrôles de la validité veillent à ce que la réponse et son format soient appropriés. Ils peuvent consister à comparer la réponse à une liste de réponses acceptables (p. ex. une liste de codes de diagnostic) ou simplement à vérifier que le format de la réponse est approprié (p. ex. un code de quatre caractères). Ainsi, le contrôle de la validité d'un élément de données relatif à une date sert à vérifier que la réponse respecte un format de date acceptable (p. ex. AAAAMMJJ) ou que la réponse est une date acceptable.

Il est important de savoir qu'une réponse valide n'est pas nécessairement correcte; cela veut simplement dire que la réponse *pourrait* être correcte. Par exemple, si une chirurgie a eu lieu le 27 février, la date est valide, mais il ne s'agit pas nécessairement de la date correcte. Toutefois, si la date de la chirurgie indiquée est le 31 février, on sait qu'il s'agit d'une date incorrecte puisqu'elle n'est pas valide.

Les données invalides d'une base de données susciteront rapidement des questions au sujet de la qualité des données. Puisqu'il est presque impossible de justifier des données invalides (comment expliquer qu'un patient pèse - 20 kg?), il est très important de les déceler. Selon la nature de la base de données, les données invalides peuvent être exclues, renvoyées au fournisseur pour correction, signalées pour imputation ou simplement signalées comme invalides et traitées séparément.

Ce critère est respecté si tous les éléments de données recueillis font l'objet de contrôles de la validité et si toutes les données invalides sont signalées.

**Critère 19** *Les règles de vérification et l'imputation sont logiques et appliquées de façon uniforme.*

Les règles de vérification sont logiques si elles ont du sens par rapport aux données recueillies. L'imputation ne devrait pas servir à modifier des données qui *pourraient* être correctes, mais plutôt à modifier des données qui sont de toute évidence incorrectes. Il importe de noter que le contrôle de certaines vérifications mises en œuvre à une étape (soit lors de la saisie des données ou du traitement des données) n'entre pas en conflit avec une autre vérification mise en œuvre à une autre étape.

Ce critère est respecté si les règles de vérification et l'imputation sont jugées logiques et si elles sont appliquées de façon uniforme afin que les données fournies soient le plus utiles possible.

**Critère 20** *Les rapports de vérification destinés aux utilisateurs sont faciles à utiliser et à comprendre.*

Les rapports de vérification doivent indiquer clairement les unités ou les éléments de données qui ont satisfait ou n'ont pas satisfait aux vérifications ainsi que la raison de la non-satisfaction.

Pour que le processus de vérification soit efficace, il est essentiel que les rapports de vérification soient faciles à comprendre. Si les données sont retournées au fournisseur de données pour modification, il est particulièrement important de lui indiquer clairement la raison pour laquelle les unités ou les éléments de données n'ont pas satisfait aux vérifications. Si la raison n'est pas précisée, le fournisseur de données risque de consacrer beaucoup de temps à un examen manuel afin de déterminer les modifications à apporter.

Le nombre de fois qu'une règle de vérification donnée est appliquée a des répercussions sur la qualité des données. Le fait qu'une règle de vérification soit appliquée plus qu'elle ne devrait l'être peut signifier qu'elle est trop restrictive ou que la qualité des données entrantes est exceptionnellement mauvaise. Dans tous les cas, les données et la règle de vérification doivent être examinées et le fournisseur pourrait devoir soumettre de nouveau les données (si possible). Enfin, les rapports de vérification destinés aux utilisateurs doivent faire clairement état de la période visée (p. ex. date et référence temporelle).

Ce critère est respecté si les rapports de vérification sont conviviaux et faciles à comprendre.

**Critère 21** *Le processus d'imputation est automatisé et conforme aux règles de vérification.*

Comme il a été défini ci-dessus, l'imputation consiste à déterminer et à attribuer des valeurs de remplacement au besoin pour les données manquantes, invalides ou incohérentes. Lorsqu'un élément de données non applicable est manquant, le recours à un indicateur dans ce champ ou dans un autre champ signifiant que l'élément de données n'est pas applicable ne doit pas être considéré comme une imputation.

L'imputation est une pratique statistique acceptée lorsqu'elle est justifiée et exécutée de façon adéquate. Une méthode d'imputation entièrement testée, approuvée et automatisée est considérée comme appropriée. En effet, l'imputation manuelle est souvent subjective, difficile à retracer, impossible à reproduire et facilement mise en doute (pourquoi cette valeur plutôt qu'une autre?). L'imputation automatisée appropriée est moins subjective, facile à retracer et plus facile à justifier lorsqu'elle est effectuée correctement. En outre, les valeurs imputées devraient passer les vérifications qui ont été appliquées aux données déclarées ou recueillies.

L'imputation peut avoir des conséquences énormes sur la qualité des données si elle est mal exécutée. C'est pourquoi il est recommandé d'élaborer tout système d'imputation avec l'aide d'un statisticien ou d'un spécialiste de la méthodologie.

Ce critère est respecté si le processus d'imputation est automatisé et conforme aux règles de vérification.

## Traitement et estimation

### Critères

- 22 *La documentation sur toutes les activités de traitement des données est tenue à jour.*
- 23 *Les spécifications techniques pour la banque de données sont tenues à jour.*
- 24 *Les modifications à la structure sous-jacente ou aux programmes de traitement ou d'estimation d'une banque de données ont été testées.*
- 25 *Les données brutes, conformément à la politique de l'ICIS en matière de maintien des données, sont sauvegardées dans un endroit sécuritaire.*
- 26a *Les statistiques agrégées d'une banque de données ont été comparées, dans la mesure du possible, à des statistiques similaires provenant d'une autre banque de données de l'ICIS ou d'une source externe.*
- 26b *La variance de l'estimation, comparée à l'estimation elle-même, se situe à un niveau acceptable.*

Le **traitement** correspond à l'application systématique de programmes ou de procédures à une base de données à n'importe quelle fin ou presque. L'application doit être réalisée selon un ordre logique, où chaque programme est autonome dans l'exécution de fonctions précises. Généralement, le traitement transforme les données obtenues au cours de la collecte dans un format qui se prête à l'analyse ou à la présentation en tableau. Par exemple, le traitement des données soumises à l'ICIS se compose souvent des étapes suivantes :

1. Les erreurs dans les unités ou les éléments de données sont relevées lors du processus de vérification.
2. La façon de modifier les unités ou les éléments de données est déterminée (suivi ou imputation).
3. Les corrections sont apportées aux données, le cas échéant.
4. Les éléments de données sont dérivés ou regroupés, au besoin.
5. Des fichiers sont créés aux fins d'analyse ou de présentation en tableau.

Il faut garder en tête que toute erreur se produisant lors d'une de ces étapes peut avoir des répercussions considérables sur la qualité des données d'une base de données.

Puisque des erreurs peuvent se produire à n'importe quelle étape du traitement, particulièrement lors d'activités manuelles et répétitives, le traitement doit faire l'objet d'une surveillance et des mesures correctives doivent être prises au besoin aux fins de maintien ou d'amélioration de la qualité. Pour ce faire, des procédures de contrôle et d'assurance de la qualité sont généralement mises en œuvre.

Les procédures de contrôle de la qualité permettent de vérifier les données entrantes et celles qui passent par les systèmes de traitement. Si les erreurs sont relevées au début du processus, des modifications peuvent être apportées rapidement et la qualité des données ne sera pas compromise. Il importe de souligner que dans les procédures de contrôle de la

qualité, l'accent est mis sur les données mêmes. Par exemple, les procédures de contrôle de la qualité comprennent la vérification des unités présentes en double dans la banque de données avant l'ajout de nouvelles unités et le suivi des rejets à la vérification, une mesure qui permet de s'assurer que les données sont correctement soumises de nouveau.

Les procédures d'assurance de la qualité, quant à elles, sont axées sur le processus. Elles comportent des procédures de contrôle de la qualité ainsi qu'un certain nombre d'autres procédures visant à atteindre le niveau souhaité de qualité des données et des estimations qui en découlent. Le calcul de statistiques qui évaluent la qualité, la formation adéquate des personnes responsables du traitement des données, l'utilisation de données d'essai dans le but de vérifier si les nouvelles composantes ont été mises en œuvre correctement et la tenue d'une documentation appropriée sont d'autres exemples de procédures faisant partie de l'assurance de la qualité.

L'**estimation** est le regroupement de données, de quelque façon que ce soit, destiné à produire une valeur qui représente la population de référence et à tirer des conclusions sur cette population. Elle peut être réalisée sur des données tirées d'un échantillon ou d'un recensement, des données dont le taux de réponse est inférieur à 100 % ou des données présentant des problèmes de couverture. Les données sont ensuite ajustées, au besoin, de façon à ce que les valeurs calculées représentent la population de référence. Il est important de noter que presque toutes les valeurs produites à partir des bases de données (même les totaux agrégés) sont des estimations, dans le sens où elles représentent des approximations de la réalité et non pas nécessairement la valeur réelle puisqu'elles comportent souvent un certain type d'erreur.

Les données peuvent présenter différents types d'erreurs. Ci-dessous sont décrits cinq types d'erreurs ainsi que la façon de les mesurer et de les déclarer. Quatre sont des erreurs non liées à l'échantillonnage et l'autre est une erreur d'échantillonnage.

#### **Erreurs non liées à l'échantillonnage**

1. Erreur de couverture (décrite à la section Couverture), mesurée au moyen des taux de sous-dénombrement et de surdénombrement.
2. Erreur de non-réponse, mesurée au moyen des taux de non-réponse ou de réponse. Les sections Non-réponse totale et Non-réponse partielle (ou non-réponse par question) expliquent plus en détail ce type d'erreur.
3. Erreur de mesure (décrite à la section Erreur de mesure), mesurée au moyen des taux de divergence.
4. Erreur de traitement (décrite à la section Vérification et imputation), déclarée au moyen des taux de rejets à la vérification et des taux d'imputation.

**NOUVEAU**

## Erreur d'échantillonnage

Les erreurs d'échantillonnage se produisent lorsque seule une partie de la population est sélectionnée (ou sondée). Si l'échantillon sélectionné est représentatif de la population de référence et que les erreurs non liées à l'échantillonnage sont minimales, l'estimation découlant de l'enquête par sondage avec l'erreur d'échantillonnage associée devrait, dans la majorité des cas, comprendre l'estimation réelle (celle qui aurait été établie si les réponses de l'ensemble des unités de la base de sondage avaient été obtenues).

Les trois mesures courantes de l'erreur d'échantillonnage sont la *variance*, l'*erreur-type* et le *coefficient de variation*.

1. La variance est la mesure de la variabilité des estimations obtenues à partir du prélèvement de tous les échantillons possibles de la population de référence. Cette mesure n'est souvent pas déclarée lorsqu'il s'agit d'un très grand nombre. L'erreur-type ou le coefficient de variation sont les statistiques habituellement déclarées.
2. L'erreur-type correspond à la racine carrée de la variance.
3. Le coefficient de variation est la mesure absolue de la dispersion. Elle est obtenue lorsqu'on exprime l'erreur-type en pourcentage de l'estimation, comme ceci :

$$\frac{\text{Erreur-type} \times 100}{\text{Estimation}}$$

Il est fortement recommandé que le calcul de ces mesures et la conception d'une enquête par sondage soient réalisés en consultation avec le personnel de la Division de la qualité des données.

Voici un exemple de ces trois mesures d'erreur d'échantillonnage. On a demandé une estimation du nombre total de personnes ayant reçu un diagnostic de diabète dans une province donnée au moyen d'une enquête par sondage. Il a été déterminé qu'un échantillon représentatif compterait 2 000 personnes. Selon les réponses obtenues auprès de cet échantillon, le nombre de personnes diabétiques dans cette province a été estimé à 10 000. La variance de cette estimation a été établie à 250 000 (habituellement, la mesure de la variance n'est pas incluse dans le rapport puisque les autres mesures suffisent). Une fois la variance connue, il est facile de calculer l'erreur-type et le coefficient de variation. L'intervalle de confiance de l'estimation est souvent fixé à 95 %. Ainsi, le coefficient applicable à utiliser d'après une table des probabilités normales est de 1,96. En s'appuyant sur cette table, il est possible de déterminer le coefficient applicable pour tout intervalle de confiance. Une fois le coefficient déterminé, l'intervalle de confiance est calculé selon la formule suivante :

$$\begin{aligned} \text{Intervalle de confiance (limite inférieure)} &= \text{estimation} - \text{coefficient} \times \text{erreur-type} \\ \text{Intervalle de confiance (limite supérieure)} &= \text{estimation} + \text{coefficient} \times \text{erreur-type} \end{aligned}$$

Le tableau suivant résume les mesures estimées relatives à l'exemple ci-dessus.

<b>Estimation du nombre total</b>	10 000 personnes diabétiques
<b>Variance de l'estimation</b>	250 000
<b>Erreur-type</b>	Racine carrée de 250 000 = 500
<b>Coefficient de variation</b>	$(500/10\ 000) \times 100 = 5 \%$
<b>Intervalle de confiance de 95 %</b>	$10\ 000 \pm 1,96 \times 500 = 10\ 000 \pm 980$ = [9 020; 10 980]

L'intervalle de confiance de 95 % signifie qu'il y a 95 chances sur 100 que la valeur réelle du nombre total de personnes diabétiques dans cette province se situe entre 9 020 et 10 980.

**Critère 22** *La documentation sur toutes les activités de traitement des données est tenue à jour.*

Le traitement correspond à la séquence des étapes utilisées notamment pour le chargement des données, la vérification des données et la production d'estimations. Le départ de personnel de l'unité de travail ou de l'organisme peut facilement entraîner une perte de connaissances relatives aux étapes de traitement et ainsi causer des erreurs dans les données. La documentation relative à toutes les activités de traitement des données devrait se trouver au même endroit et un endroit unique pour chaque processus suffit. Les étapes doivent être suffisamment documentées pour qu'un nouveau venu dans l'équipe de projet puisse se servir des documents pour suivre et mettre en œuvre les processus.

La meilleure façon de mettre en œuvre un ensemble complet de documents est de se conformer aux composantes proposées dans la documentation sur les métadonnées (section 4.3).

Ce critère est respecté si tous les processus **utilisés par le personnel de la banque de données** sont documentés de façon appropriée.

**Critère 23** *Les spécifications techniques pour la banque de données sont tenues à jour.*

La façon dont les systèmes et les applications d'une banque de données sont conçus peut influencer sur la qualité des données de cette banque. Il est donc important que les spécifications soient mises en œuvre comme prévu et qu'elles soient toutes mises à l'essai. Il est tout aussi important que ces spécifications techniques soient tenues à jour sous forme de documentation. La documentation permet de valider aisément les systèmes, les programmes et les applications et facilite la mise en œuvre de modifications à apporter, s'il y a lieu. Une bonne documentation devrait être accessible et facile à comprendre par un nouveau membre de l'équipe du projet. De plus, cette documentation devrait être revue régulièrement et comprendre les mises à jour des spécifications techniques.

Ce critère est respecté si la documentation sur les systèmes, les programmes ou les applications de la banque de données est tenue à jour.

**Critère 24** *Les modifications à la structure sous-jacente ou aux programmes de traitement ou d'estimation d'une banque de données ont été testées.*

Même si des révisions sont parfois nécessaires à la suite de modifications apportées à une banque de données, les modifications de programmes peuvent avoir des conséquences inattendues. Par exemple, les modifications apportées aux éléments de données peuvent nécessiter une mise à l'essai considérable, qu'il s'agisse d'un changement au nom de l'élément de données ou de l'ajout ou de la suppression d'un élément de données. Il est nécessaire de vérifier les programmes modifiés pour s'assurer que les résultats attendus sont atteints. Il est également prudent d'examiner les effets des changements en aval : par exemple, remplacer le format numérique d'un élément de données par un format composé de caractères peut influencer sur les programmes qui traitent ultérieurement l'élément de données comme étant numérique. Des tests d'acceptation des unités, des systèmes et des utilisateurs devraient être réalisés en vue de prévenir les résultats inattendus dans un environnement de production.

Ce critère est respecté si les systèmes sont testés lorsque des modifications ont lieu. Si aucune révision n'a été effectuée au cours de la dernière année, cette cote n'est pas applicable.

**Critère 25** *Les données brutes, conformément à la politique de l'ICIS en matière de maintien des données, sont sauvegardées dans un endroit sécuritaire.*

Comme des vérifications sont nécessaires et que des erreurs peuvent se produire lors du traitement, il est important que les données brutes, telles qu'elles sont fournies par les fournisseurs de données, soient sauvegardées dans un endroit sécuritaire. Les données devraient être sauvegardées de façon à ce qu'elles ne puissent pas être modifiées ni supprimées accidentellement. Le personnel de la base de données peut ainsi consulter les données originales, non modifiées. Lorsque la politique de l'ICIS en matière de maintien des données est respectée, il est beaucoup plus facile de faire face aux différentes situations. Par exemple, les erreurs qui se sont produites lors du traitement des données peuvent être repérées, quantifiées et réduites au maximum au moyen d'un traitement subséquent à l'aide des données brutes. Par ailleurs, les problèmes soulevés au sujet des données peuvent être résolus à partir des données brutes. Si les résultats d'une analyse sont douteux, les données brutes peuvent s'avérer très utiles dans la vérification de l'analyse originale grâce à la répétition des étapes de traitement et d'analyse.

Il faut savoir que si les données brutes ne peuvent être sauvegardées ou ne l'ont pas été, le personnel de la section doit, à tout le moins, être en mesure de récupérer les données à partir des étapes précédentes du traitement. Les étapes du traitement pourront ainsi être répétées.

Ce critère est respecté si les données soumises par les fournisseurs sont sauvegardées dans un endroit sécuritaire ou si les changements apportés aux données peuvent être reproduits.

NOUVEAU

**Critère 26a** *Les statistiques agrégées d'une banque de données ont été comparées, dans la mesure du possible, à des statistiques similaires provenant d'une autre banque de données de l'ICIS ou d'une source externe.*

La comparaison des statistiques agrégées d'une banque de données à une autre permet d'obtenir une indication de la qualité des données. Cette comparaison devrait être réalisée lorsque des statistiques agrégées ont été produites pour le même élément de données au sein de la même population de référence et de la même période. Des différences pourraient indiquer que le processus de regroupement présente un problème. Les statistiques agrégées de banques de données dont les données sont fondées sur un recensement et celles dont les données sont fondées sur des échantillons aléatoires peuvent être comparées.

Ce critère est respecté si les statistiques agrégées d'une banque de données ont été comparées, dans la mesure du possible, à celles de banques de données similaires.

NOUVEAU

**Critère 26b** *La variance de l'estimation, comparée à l'estimation elle-même, se situe à un niveau acceptable.*

Ce critère s'applique uniquement aux estimations fondées sur un échantillon aléatoire dont la probabilité de sélection est connue. **Les bases de données qui n'utilisent pas ce type d'échantillons devraient coter ce critère comme *non applicable*.** Afin d'éviter de confondre les termes, il est important que les bases de données fondées sur des données de recensement n'emploient pas le terme « variance » pour décrire la variabilité de leurs données. Elles doivent plutôt se servir d'autres mesures de la variabilité, comme l'écart, l'intervalle interquartile, l'écart moyen ou l'écart absolu moyen.

La **variance** d'une estimation est une mesure de la variabilité des estimations obtenues au moyen du prélèvement de tous les échantillons possibles de la population de référence. Le niveau acceptable de variance dépend de la valeur de l'estimation découlant de l'échantillon. Au moment de formuler des énoncés au sujet de la variabilité d'une estimation, il est important de déclarer l'erreur-type ou le coefficient de variation. Le coefficient de variation, exprimé en pourcentage, correspond à l'erreur-type divisée par la valeur de l'estimation. Il s'agit de la statistique sur laquelle se fonde le tableau d'attribution des cotes ci-dessous, dont l'objectif est de guider les personnes qui publient des estimations fondées sur des données d'échantillon. Selon la valeur du coefficient de variation pour une estimation, ce tableau peut servir à établir si une estimation se situe à un niveau acceptable aux fins de publication, si elle doit être accompagnée d'un avertissement à l'intention des utilisateurs ou si elle ne devrait pas être publiée en raison de son manque de fiabilité.

**Tableau D Guide de cotation relatif aux coefficients de variation estimés**

Coefficient de variation (pourcentage)	Ligne directrice
Moins de 16,6 %	Niveau acceptable
De 16,6 à 33,3 %	Procéder avec prudence
Plus de 33,3 %	Ne pas publier

Le tableau D se veut une ligne directrice qui permet de déterminer si l'erreur d'échantillonnage se situe à un niveau acceptable ou non. Prenez note que le coefficient de variation obtenu dépend grandement de la valeur de l'estimation puisque cette dernière correspond au dénominateur dans le calcul du coefficient. La taille de l'échantillon relativement au nombre d'unités de la base de sondage dans la population peut également influencer sur la variance. En général, dans le cas de coefficients de variation d'estimations se situant à un niveau acceptable (moins de 16,6 %), la cote *respecté* devrait être attribuée à ce critère.

En revanche, le critère devrait recevoir la cote *non respecté* si le coefficient de variation est supérieur à 16,6 % ou si le niveau n'est pas acceptable pour les clients de la banque de données. Cela dit, la décision finale à savoir si la variance est acceptable ou non dépend de l'utilisateur et des besoins en matière de données.

Ce critère est respecté si la variance de l'estimation, comparée à l'estimation elle-même, se situe à un niveau acceptable pour les utilisateurs des données.

## 2 Dimension de l'actualité

L'actualité désigne principalement le degré d'actualité des données au moment de leur diffusion. Le caractère actuel des données est mesuré d'après l'intervalle entre la fin de la période de référence à laquelle les données se rapportent et la date à laquelle les données ont été mises à la disposition des utilisateurs. L'actualité est donc étroitement liée à la pertinence, car si le délai est trop long, les données pourraient ne plus être pertinentes pour les besoins des utilisateurs. Bien que la production des données dans les délais soit nécessaire pour en assurer la pertinence, les calendriers acceptables peuvent varier selon les diverses banques de données de l'ICIS. On ne peut astreindre aux mêmes calendriers les bases de données qui dépendent des données provenant d'autres bases de données.

Une trop grande importance accordée à la dimension de l'actualité risque de compromettre l'exactitude. Par exemple, si les membres du personnel d'un hôpital ou d'une clinique privée n'ont pas suffisamment de temps pour remplir un sondage sur leurs équipements d'imagerie médicale, les statistiques sur ces équipements peuvent être opportunes mais incomplètes ou inexactes. Cette perte d'intégralité ou d'exactitude ne vaut pas le gain d'actualité. Il faut aussi prévoir suffisamment de temps entre la fermeture annuelle de la banque de données et la diffusion des données afin de vérifier ces dernières et de documenter les limites pour les utilisateurs. Comme il est toujours possible d'accroître le contrôle de la qualité et la documentation, il est nécessaire d'atteindre un équilibre entre l'actualité et l'exactitude. À tout le moins, la documentation recommandée sur la qualité des données doit être disponible au moment de la diffusion. Veuillez consulter la présentation hiérarchique des dimensions à la section 3.3 pour mieux comprendre le lien entre l'actualité et les autres dimensions de la qualité des données.

Le but de la dimension de l'actualité est d'évaluer le caractère actuel des données et la disponibilité de la documentation recommandée sur la qualité des données au moment de la diffusion. Il importe de surveiller l'actualité pour relever les bonnes pratiques à l'ICIS de même que les cas de graves retards. Les critères de cette dimension visent aussi à évaluer si les principaux rapports de la banque de données ont été diffusés au moment prévu. La dimension comporte les caractéristiques suivantes :

- Actualité des données au moment de leur diffusion (a-t-on rendu les données disponibles dans un délai raisonnable?)
- Actualité de la documentation (a-t-on diffusé les principaux documents à temps?)

Dimension	Caractéristiques	Critères
Actualité	Actualité des données au moment de leur diffusion	27 à 30
	Actualité de la documentation	31 et 32

## Actualité des données au moment de leur diffusion

### Critères

- 27 *L'écart entre la date réelle de diffusion et la fin de la période de référence est raisonnablement bref.*
- 28 *La date officielle de diffusion a été annoncée avant la diffusion.*
- 29 *La diffusion a eu lieu à la date officielle prévue.*
- 30 *Les activités de traitement des données sont revues régulièrement aux fins d'amélioration de l'actualité des données.*

Cette caractéristique évalue d'abord et avant tout l'actualité des données à l'intérieur d'une banque de données au moment de leur diffusion. Le caractère à jour des données, la composante fondamentale de l'actualité, est mesuré d'après l'intervalle entre la date de diffusion et la dernière date à laquelle les données se rapportent. L'intervalle doit être suffisamment court pour que les données soient encore conformes à leurs principales fins. Deux autres aspects s'appliquent également : si les données sont diffusées à temps et si les méthodes de la banque de données sont aussi efficaces que possible. Si les méthodes utilisées pour traiter et analyser les données sont aussi exactes et efficaces que possible, la diffusion de données ne sera pas inutilement retardée.

**Critère 27** *L'écart entre la date réelle de diffusion et la fin de la période de référence est raisonnablement bref.*

La **date de diffusion** correspond à la date officielle à laquelle une diffusion annuelle de données provenant d'une banque de données devient disponible pour les utilisateurs. On entend par **période de référence** la période à laquelle les données se rapportent réellement. Le début de la période de référence est la première date à laquelle les données se rapportent et la fin de la période de référence est la dernière date à laquelle les données se rapportent. Dans le cas des banques de données qui n'ont pas de diffusion annuelle de données, toute diffusion de données aux principaux intervenants devrait servir de point de comparaison. Ce critère ne s'applique pas aux bases de données de nature longitudinale ni à celles qui n'ont pas de fin de période de référence.

Les banques de données ont des normes différentes pour définir ce qui est raisonnablement bref. En règle générale, pour les diffusions annuelles de données, un délai de 6 à 9 mois entre la fin de la période de référence et la date de diffusion est considéré comme souhaitable, et un délai allant jusqu'à 12 mois est considéré comme acceptable.

Ce critère est respecté si le temps écoulé entre la date de diffusion et la fin de la période de référence est raisonnablement bref.

**Critère 28** *La date officielle de diffusion a été annoncée avant la diffusion.*

Les dates officielles des diffusions aux principaux intervenants, comme la diffusion annuelle de données, devraient être annoncées longtemps à l'avance. Cette annonce devrait être faite aux principaux utilisateurs des données, à l'interne ou à l'externe. Les diffusions peuvent être annoncées dans le calendrier des produits livrables, dans le Guide des produits et services de l'ICIS, sur le site Web de l'ICIS ou dans des avis envoyés aux principaux intervenants. Il est important pour les utilisateurs de connaître ces dates, car ils peuvent à leur tour élaborer leurs propres plans opérationnels. Des dates réalistes devraient être choisies pour chaque diffusion. Si elles ne répondent pas aux besoins des clients, il faudrait envisager des solutions de rechange.

Ce critère est respecté si la date officielle de la diffusion annuelle des données ou de la diffusion de données aux principaux intervenants a été planifiée et annoncée au moins six mois à l'avance.

**Critère 29** *La diffusion a eu lieu à la date officielle prévue.*

Il est important pour les utilisateurs que les données soient diffusées à temps et comme prévu. L'intervalle entre la date réelle et la date prévue de la diffusion peut avoir des répercussions sur le cycle de production de ceux qui dépendent des données. Il est recommandé de vérifier le respect des dates de diffusion annoncées préalablement, les modifications apportées à ces dates et les raisons qui justifient ces modifications. Pour diffuser des données plus rapidement sans nuire aux autres dimensions de la qualité, il est nécessaire que le calendrier de diffusion des données soit ajusté en fonction des expériences antérieures. De plus, les sections de l'ICIS devraient toujours s'efforcer de réduire le cycle de production. Pour ce faire, elles peuvent mettre au point des méthodes permettant de traiter les données plus rapidement.

Ce critère est respecté si les données ont été diffusées à la date officielle de diffusion ou avant. En ce qui concerne les diffusions de données aux principaux intervenants, le critère repose sur la comparaison entre la date de diffusion prévue (indiquée dans le calendrier des produits livrables) et la date de diffusion réelle.

**Critère 30** *Les activités de traitement des données sont revues régulièrement pour en améliorer l'actualité.*

Les programmes et les systèmes utilisés pour préparer et analyser les données devraient être revus continuellement afin qu'ils soient le plus efficaces possible en matière de production de données en temps opportun. Par exemple, plusieurs programmes peuvent être combinés en vue de réduire le nombre d'entrées manuelles ainsi que le temps requis pour la gestion et l'analyse des données et la création de rapports. La comparaison des méthodes des banques de données à celles d'autres banques de données similaires externes ou internes peut donner lieu à l'amélioration de l'efficacité. Il est aussi possible de revoir les méthodes existantes à la lumière des nouvelles technologies, procédures ou pratiques normalisées des banques de données susceptibles d'être plus efficaces (comme le Service de soumission électronique des données ou eDSS) et ainsi produire des données plus exactes. De nouvelles méthodes ou technologies, à condition d'avoir été rigoureusement testées, peuvent constituer le meilleur moyen d'améliorer à la fois l'actualité et l'exactitude des données.

Ce critère est respecté si les activités de traitement des données de la banque de données ont été revues au moins une fois dans l'année précédente.

## Actualité de la documentation

### Critères

*31 La documentation recommandée sur la qualité des données était disponible au moment de la diffusion des données ou des rapports.*

*32 Les principaux rapports ont été diffusés à la date prévue.*

Cette caractéristique évalue si les principaux documents ont été rendus disponibles à temps. Plus précisément, elle évalue si la documentation recommandée sur la qualité des données et les principaux rapports des banques de données ont été rendus disponibles au moment opportun ou prévu. La documentation sur la qualité des données a pour objectif d'informer les utilisateurs des principales limites associées aux données afin qu'ils puissent déterminer si les données conviennent à l'utilisation qu'ils désirent en faire. Ce type d'information est aussi nécessaire pour interpréter correctement les résultats tirés de ces données. Il est donc essentiel que la documentation sur la qualité des données accompagne toute diffusion de données ou tout rapport important.

**Critère 31** *La documentation recommandée sur la qualité des données était disponible au moment de la diffusion des données ou des rapports.*

Il est important que la documentation sur la qualité des données soit rendue disponible au moment où les utilisateurs ont accès, tant à l'interne qu'à l'externe, aux données ou aux rapports résumant les données. Il faut donc allouer suffisamment de temps entre la fermeture de la base de données et la diffusion des données. En ce qui concerne les bases de données longitudinales, la documentation sur la qualité des données doit être disponible au moment où les données sont diffusées ou peuvent être consultées, ou encore lorsqu'un rapport est rédigé. La documentation sur la qualité des données inclut notamment des renseignements sur le contexte, les sources de données et la méthodologie, les variables

et les concepts mesurés, l'exactitude des données, la comparabilité des données, les publications et les produits, ainsi que les services à utiliser pour obtenir de plus amples renseignements. La section 4 présente plus en détail les composantes qui devraient être incluses dans toute documentation sur la qualité des données.

Ce critère est respecté si la documentation sur la qualité des données était disponible au moment de la diffusion des données ou des rapports.

**Critère 32** *Les principaux rapports ont été diffusés à la date prévue.*

Il est important d'annoncer à l'avance les dates de diffusion des principaux rapports et de respecter ces dates. Le non-respect des dates de diffusion des rapports écrits représente non seulement un désagrément pour les utilisateurs qui ont planifié certaines activités en fonction de la date de diffusion prévue, mais il peut aussi miner leur confiance.

Ce critère est respecté si les principaux rapports de la banque de données ont été diffusés à la date prévue.

### 3 Dimension de la comparabilité

La comparabilité représente le niveau d'uniformité des banques de données au fil du temps et leur utilisation des conventions normalisées (comme les éléments de données ou les périodes de déclaration), caractéristiques qui déterminent leur similitude avec les autres bases de données. Au sein d'un organisme comme l'ICIS, qui gère de nombreuses banques de données, la comparabilité facilite la compréhension, l'interprétation et la mise à jour des données. Elle est aussi directement liée à la partie du mandat de l'ICIS qui touche le développement et la mise à jour d'un système exhaustif et intégré d'information sur la santé. Il est nécessaire d'utiliser des populations de référence identiques ou similaires pour établir des comparaisons entre différentes banques de données. Les bases de données comparables utilisent les mêmes définitions de données, recueillent le même type de données et peuvent être couplées à d'autres bases de données similaires, d'où la possibilité de combiner des données de différentes sources afin de répondre à d'importantes questions de recherche qui ne pourraient autrement être étudiées. La recherche sur la continuité des soins en est un très bel exemple, compte tenu de la variété des bases de données cliniques requises pour l'analyse (des soins d'urgence aux soins aux malades chroniques).

La comparabilité a aussi l'avantage de permettre l'évaluation d'autres aspects de la qualité des données, comme l'exactitude. La comparaison de bases de données similaires peut être un moyen efficace d'examiner les questions associées à la couverture, aux erreurs de codification et à la non-réponse.

La dimension de la comparabilité permet d'évaluer à quel point les bases de données répondent à une norme commune. Elle comporte les caractéristiques suivantes :

- Normes du Dictionnaire de données (la base de données utilise-t-elle les normes de l'ICIS pour les définitions de données?)
- Normalisation (peut-on tirer des groupements communs à partir des données?)
- Couplage (peut-on unir des bases de données par un élément de données commun?)
- Équivalence (les valeurs des données sont-elles converties correctement?)
- Comparabilité rétrospective (les données sont-elles comparables au fil du temps?)

Dimension	Caractéristiques	Critères
Comparabilité	Normes du Dictionnaire de données	33 et 34
	Normalisation	35 et 36
	Couplage	37 à 40
	Équivalence	41 et 42
	Comparabilité rétrospective	43 à 45

## Normes du Dictionnaire de données

### Critères

*33 Tous les éléments de données font l'objet d'une évaluation visant à déterminer s'ils sont inclus dans le Dictionnaire de données de l'ICIS.*

*34 Les éléments de données d'une banque de données qui sont inclus dans le Dictionnaire de données de l'ICIS doivent être conformes aux normes de ce Dictionnaire.*

Cette caractéristique porte sur les éléments de données de la base de données et leur conformité au Dictionnaire de données de l'ICIS, qui contient les éléments et les définitions approuvés par l'équipe interne du Dictionnaire. Le but est d'appliquer les mêmes définitions à toutes les bases de données pour des éléments de données communs et ainsi d'éliminer la confusion chez les fournisseurs de données et les chercheurs.

Les normes du Dictionnaire de données sont actuellement revues et révisées. Au moment de la dernière révision du présent cadre de la qualité des données, tous les éléments de données se trouvant sous la classe d'objets Dispensateur de soins de santé avaient été finalisés. Veuillez consulter la page [www.icis.ca/dictionnairededonnees](http://www.icis.ca/dictionnairededonnees) pour connaître les dernières définitions des éléments de données finaux. Bien que l'adoption des normes du Dictionnaire soit obligatoire pour les nouvelles banques de données (ou celles qui sont remaniées), il n'est pas nécessaire pour l'instant d'appliquer les normes aux banques de données existantes. Il est toutefois recommandé d'assurer la comparabilité d'une banque de données à l'autre.

**Critère 33** *Tous les éléments de données font l'objet d'une évaluation visant à déterminer s'ils sont inclus dans le Dictionnaire de données de l'ICIS.*

Le Dictionnaire de données de l'ICIS constitue la norme en matière d'éléments de données que toutes les banques de données à l'ICIS devraient suivre. Comme le Dictionnaire ne cesse d'évoluer, il contient des définitions mises à jour pour bien des éléments de données relatifs aux dispensateurs de soins de santé. Par conséquent, il importe que tous les éléments de données des banques de données existantes et de celles en cours de développement soient revus périodiquement par rapport au Dictionnaire de données de l'ICIS (voir [www.icis.ca/dictionnairededonnees](http://www.icis.ca/dictionnairededonnees)). L'évaluation déterminera si les normes utilisées pour les éléments de données des banques de données correspondent à celles du Dictionnaire de données de l'ICIS. Les éléments qui ne correspondent pas aux normes du Dictionnaire doivent être relevés et éventuellement modifiés. De plus, les éléments de données de la banque de données auxquels ne correspond aucune norme dans le Dictionnaire de données de l'ICIS devraient être signalés.

Ce critère est respecté si la base de données a été évaluée par rapport au Dictionnaire de données de l'ICIS au moins une fois dans l'année précédente.

**Critère 34** *Les éléments de données d'une banque de données qui sont inclus dans le Dictionnaire de données de l'ICIS doivent être conformes aux normes de ce Dictionnaire.*

Tous les éléments de données d'une banque de données qui sont communs aux éléments de données finaux du Dictionnaire de données de l'ICIS (voir [www.icis.ca/dictionnairededonnees](http://www.icis.ca/dictionnairededonnees)) devraient partager les mêmes attributs. Il y a plusieurs facteurs à considérer dans l'évaluation de la conformité, comme le nom de l'élément de données, le domaine des valeurs, le type de données et la longueur de l'enregistrement. Idéalement, tous les attributs devraient être identiques. Lorsque ce n'est pas le cas, la conformité peut être partielle ou inexistante. Par exemple, un élément de données représentant les unités relatives au sexe d'un dispensateur de soins de santé devrait avoir les attributs suivants :

Attribut	Description de l'attribut
Nom	Sexe du dispensateur
Type de données	Caractère
Longueur maximale	1
Domaine de valeur	M = Homme F = Femme U = Non différencié 7 = Non recueilli 8 = Sans objet 9 = Inconnu

Il y a correspondance partielle lorsqu'une partie seulement des attributs sont les mêmes (comme la longueur et le domaine) et il y a correspondance parfaite lorsque toutes les caractéristiques sont identiques. En général, les différences dans les domaines de valeur sont plus graves puisque chaque enregistrement doit être modifié pour qu'il soit conforme à une norme donnée.

**Remarque :** Le Dictionnaire de données de l'ICIS n'est pas complet pour tous les types d'éléments de données. Pour cette raison, l'évaluation de la conformité ne s'applique qu'aux éléments figurant actuellement dans le Dictionnaire. Toute dérogation justifiable aux normes de l'ICIS devrait être décrite.

Ce critère est respecté s'il y a correspondance parfaite entre tous les éléments de données communs à la banque de données et au Dictionnaire de données de l'ICIS.

## Normalisation

### Critères

35 *Les données sont recueillies au niveau de détail le plus élevé possible.*

36 *Pour tout élément de données dérivé, l'élément de données original demeure accessible.*

Les bases de données groupent souvent les éléments de données de diverses façons, selon l'application ou le contexte. Toutefois, s'il devient nécessaire de comparer des données de différentes bases de données, un groupement commun doit être dérivé. Même s'il n'est ni pratique ni raisonnable de s'attendre à ce que les autres bases de données maintiennent les mêmes groupements, la saisie des données à un niveau de détail suffisamment élevé peut garantir la comparabilité. Par exemple, si l'âge d'un patient est habituellement déclaré selon des catégories d'âge de 10 ans, l'âge en années devrait aussi être disponible afin de permettre la création d'autres catégories d'âge, au besoin. De cette façon, il est possible de normaliser les éléments de données de différentes bases de données. Il est à noter que le Dictionnaire de données contient une norme pour la date de naissance, l'année de naissance et le mois de naissance du dispensateur, mais aucune norme pour l'âge du dispensateur. Par conséquent, il est possible de dériver l'âge du dispensateur à partir de n'importe laquelle de ces normes du Dictionnaire de données.

**Critère 35** *Les données sont recueillies au niveau de détail le plus élevé possible.*

Un niveau de détail élevé dans les définitions de données est important parce qu'il laisse place, dans une certaine mesure, à la conformité à différentes normes. Il est important de noter que le niveau de détail dépend des utilisateurs de la banque de données et, dans certains cas, du fournisseur de données. Par exemple, dans le cas des soins de courte durée, il est habituellement suffisant de mesurer la durée du séjour en jours. Par contre, le temps d'attente à l'urgence devrait être mesuré en minutes. La collecte de données au moyen de la date de naissance d'une personne plutôt que de son groupe d'âge, même si l'information est déclarée dans certaines catégories de groupe d'âge, en est un autre exemple. Un simple groupement dans différents groupes d'âge est possible si l'âge a été déclaré en années. Bref, un niveau de détail plus élevé n'est peut-être pas toujours nécessaire pour les utilisations courantes, mais il peut l'être pour créer de nouveaux groupements.

Ce critère est respecté si tous les éléments de données de base sont recueillis avec le niveau de détail nécessaire aux fins de publication, de couplage ou de comparaison. Toute exception à la collecte de données sans le niveau de détail le plus élevé devra être justifiée.

**Critère 36** *Pour tout élément de données dérivé, l'élément de données original demeure accessible.*

En règle générale, les éléments de données utilisés pour la création d'un autre élément de données doivent être conservés si jamais des modifications ou de nouveaux calculs devaient être effectués. Il peut s'avérer nécessaire de chiffrer les renseignements de nature délicate, comme le numéro d'assurance-maladie ou la date de naissance, ou d'en restreindre l'accès, mais ils ne devraient jamais être complètement supprimés du fichier. La simple présence de l'élément original dans le fichier de données brutes n'est pas suffisante si l'accès en est difficile. Ces éléments de nature délicate ne devraient pas être supprimés de façon permanente du fichier brut de la base de données principale; il suffit plutôt de créer un fichier de données supplémentaire comprenant uniquement les éléments de données dérivés pour tous les éléments de nature délicate qui ont été retirés ou chiffrés. Veuillez noter que ce critère s'applique aux données originales et non pas aux demandes de données qui peuvent reposer uniquement sur les éléments de données dérivés.

Ce critère est respecté seulement si les éléments de données originaux sont accessibles et s'ils n'ont pas été supprimés de façon permanente.

## Couplage

### Critères

- 37 *Les données géographiques sont recueillies au moyen de la Classification géographique type (CGT).*
- 38 *Les données sont recueillies selon le même calendrier, plus particulièrement d'une province ou d'un territoire à l'autre et au sein d'une province ou d'un territoire.*
- 39 *L'utilisation d'identificateurs permet de distinguer les établissements ou les organismes seulement aux fins de couplage rétrospectif.*
- 40 *L'utilisation d'identificateurs permet de distinguer les personnes ou les appareils seulement aux fins de couplage rétrospectif.*

On entend par **couplage** l'union d'enregistrements provenant de deux banques de données ou plus au moyen d'un ou de plusieurs éléments de données de liaison, ou encore l'union d'enregistrements à l'intérieur d'une banque de données au moyen d'un élément de données commun ou plus. Comme il existe divers types de données à l'ICIS, le couplage de données provenant de différentes sources dresse un portrait plus complet. De plus, la qualité des données peut être améliorée par le couplage, car les enregistrements en double peuvent être supprimés. **Gardez à l'esprit que les lignes directrices de l'ICIS sur la protection des renseignements personnels et la confidentialité doivent être respectées lors du couplage des données entre les banques de données. Le personnel de l'ICIS devrait consulter la page intranet du Secrétariat à la vie privée et aux services juridiques.**

Le couplage des données se déroule en trois étapes :

1. Préparation des données : suppression des doubles et conversion des fichiers de données dans le même format (SAS, Excel, etc.);
2. Choix d'un élément de données de liaison ou plus : choix d'un élément de données qui deviendra l'identificateur unique présent dans tous les fichiers couplés. Tous les éléments de données utilisés dans le couplage doivent être très fiables et exacts. Le répertoire de bénéficiaires de services doit être utilisé, s'il est disponible, pour les fichiers de données en question;
3. Évaluation de tous les enregistrements couplés : succès ou échec (l'identificateur unique ne se trouve pas dans tous les fichiers, ou un des couplages a échoué en raison d'erreurs typographiques, etc.).

**Cette caractéristique évalue si le couplage est possible et non si le couplage est effectivement réalisé.** Les critères portent sur les quatre secteurs principaux du couplage : la géographie, le temps, l'établissement et la personne ou l'appareil.

**Critère 37** *Les données géographiques sont recueillies au moyen de la Classification géographique type (CGT).*

Ce critère se rapporte à la **Classification géographique type (CGT)** telle que définie par Statistique Canada. La CGT consiste en une classification des secteurs géographiques utilisés pour la collecte et la diffusion des statistiques. On y trouve des codes de secteurs géographiques types (secteurs de recensement, divisions de recensement, subdivisions de recensement, régions métropolitaines de recensement, agglomérations de recensement et régions économiques) classés par province et territoire et par région métropolitaine. Dans ce système, différentes régions géographiques sont groupées en un système hiérarchique. Les plus petits types d'agrégation comprennent le côté d'îlot et le secteur de dénombrement. Ils sont nichés à l'intérieur de groupements de plus en plus vastes, comme le secteur de recensement et la division de recensement, pour enfin aboutir à la province et au pays. Une grande partie de l'information de nature sociale et démographique provenant du recensement est agrégée à différents niveaux de la CGT, ce qui la rend intéressante pour un vaste éventail de projets de recherche. Aux fins de couplage, il est important que les données géographiques recueillies par chaque banque de données soient conformes à la CGT. Par exemple, la saisie du code postal complet suffit, car celui-ci peut ensuite être converti vers la CGT grâce au **Fichier de conversion des codes postaux**.

Dans certains cas, l'information géographique peut s'appliquer à plus d'une entité. Les bases de données cliniques, par exemple, devraient recueillir de l'information géographique non seulement sur le patient, mais aussi sur l'établissement.

Ce critère est respecté si les entités sur lesquelles sont recueillies des données (établissements, personnes, provinces, etc.) sont identifiables soit par le code postal (les six caractères), soit par la CGT pertinente. Si le plus bas niveau géographique utilisé est la province, il faut utiliser les codes provinciaux standard de Postes Canada.

**Critère 38** *Les données sont recueillies selon le même calendrier, plus particulièrement d'une province ou d'un territoire à l'autre et au sein d'une province ou d'un territoire.*

Des périodes de référence uniformes sont importantes non seulement pour le couplage, mais aussi pour des comparaisons simples de données sommaires. Il serait curieux, par exemple, de comparer deux estimations dont l'une repose sur l'année civile et l'autre sur l'exercice financier. Il est important de recueillir des données pour une base de données selon le même calendrier au fil des ans pour éviter des lacunes dans la base de données. De plus, les dates doivent être recueillies au niveau de détail le plus élevé (AAAAMMJJ). Lorsque le format standard des dates est respecté, les données pour une année civile peuvent être obtenues même si les données ont précédemment été recueillies selon l'exercice financier.

Ce critère est respecté si l'information d'une province ou d'un territoire à l'autre et au sein d'une province ou d'un territoire est suffisante pour permettre l'analyse des données selon le même calendrier.

**Critère 39** *L'utilisation d'identificateurs permet de distinguer les établissements ou les organismes seulement aux fins de couplage rétrospectif.*

Les établissements ou les organismes représentent un niveau courant d'analyse à l'ICIS et devraient donc avoir un code d'identification unique, ce qui faciliterait le couplage rétrospectif des enregistrements. Habituellement, la province ou le territoire assigne un identificateur numérique à ses établissements ou à ses organismes. D'autres identificateurs sont aussi acceptables à condition qu'il existe une liste des changements ou un tableau des correspondances approprié, comme l'index organisationnel de l'ICIS ou encore l'Institutional Care Facility Master Inventory (ICFMI) utilisé par Statistique Canada. Si les mêmes identificateurs numériques sont utilisés entre certaines provinces, il est possible de combiner le code de la province et le numéro de l'établissement pour créer un identificateur unique aux fins de couplage. Il est à noter que le nom même de l'établissement ne convient pas au couplage. Même si les noms sont de précieux identificateurs, ils représentent tout de même des éléments de données de liaison bien limités compte tenu du manque d'homogénéité dans l'orthographe, les abréviations et le format utilisés sur différentes périodes de temps.

Ce critère est respecté si un code unique acceptable aux fins de couplage rétrospectif (identificateur assigné par la province ou le territoire, ou l'équivalent) existe pour chaque établissement ou organisme et est disponible dans la base de données.

**Critère 40** *L'utilisation d'identificateurs permet de distinguer les personnes ou les appareils seulement aux fins de couplage rétrospectif.*

Ce critère vise à assurer la présence d'un identificateur approprié qui distingue avec précision les personnes ou les appareils (c'est-à-dire les appareils médicaux) compris dans la base de données. Cet identificateur doit être unique et uniforme dans le temps et il doit pouvoir être adapté en fonction des personnes ou des appareils qui s'ajouteront à la base de données dans le futur. Pour un couplage des enregistrements plus facile, l'élément de données doit aussi être uniforme d'une base de données à l'autre. Dans le cas des bases de données cliniques, il s'agira probablement du répertoire de bénéficiaires de services ou du numéro d'assurance-maladie. Si un élément de données dépersonnalisé ou chiffré est utilisé, il devrait être possible de faire correspondre l'enregistrement au numéro d'assurance-maladie. Dans le cas des bases de données sur les dispensateurs de soins de santé ou sur le personnel, d'autres identificateurs peuvent convenir, comme ceux assignés par la province ou l'organisme de réglementation. Il est à noter que les noms ne conviennent pas au couplage, notamment en raison des risques d'erreurs typographiques et du fait que, même s'ils sont propres à une personne, plusieurs peuvent porter le même nom. En ce qui concerne les banques de données qui recueillent de l'information sur des appareils médicaux, les identificateurs comme les numéros de série et les autres numéros d'identification uniques devraient être considérés lorsqu'un appareil est constitué de plusieurs pièces possédant chacune leur propre numéro de série.

Ce critère est respecté si la base de données renferme un identificateur unique de personnes ou d'appareils susceptible d'être apparié à des enregistrements correspondants présents sur différentes périodes de temps.

## Équivalence

### Critères

*41 La méthodologie et les limites des concordances ou des conversions sont documentées.*

*42 L'ampleur des problèmes liés aux concordances et aux conversions se classe parmi l'une des catégories prédéterminées.*

On entend par équivalence la façon dont les données peuvent être mises en correspondance au fil du temps, en particulier lorsqu'il est question de systèmes de classification (p. ex. la Classification internationale des maladies [CIM], la classification anatomique, thérapeutique et chimique [ATC], les Normes SIG, etc.). Les concordances et les conversions sont simplement des tableaux utilisés pour établir ces correspondances. La mise en correspondance requise pour lier les données d'une CIM à l'autre (p. ex. CIM-10-CA à CIM-9) constitue un exemple pertinent pour l'ICIS.

Dans le cas des **conversions**, la mise en correspondance est biunivoque (un pour un). Par exemple, il est possible de créer un tableau qui met en correspondance le poids d'un patient en livres et en kilogrammes. Dans ce cas, la conversion est simple puisque la formule l'est tout autant.

Les **concordances** comportent une relation pluriunivoque (plusieurs à un ou un à plusieurs). Elles sont utilisées dans la mise en correspondance des codes de diagnostic. Par exemple, les codes de la ICD-9-CM présents dans le dossier d'un patient peuvent être mis en correspondance avec ceux de la CIM-10-CA. Dans une telle situation, un code de la ICD-9-CM pourrait être mis en correspondance avec plusieurs codes de la CIM-10-CA ou, à l'inverse, plusieurs codes de la ICD-9-CM pourraient être mis en correspondance avec un code de la CIM-10-CA.

La réussite d'une concordance ou d'une conversion s'appuie en grande partie sur la capacité à convertir des valeurs d'une classification à une autre.

**Critère 41** *La méthodologie et les limites des concordances ou des conversions sont documentées.*

En raison de la complexité de nombreuses concordances ou conversions, la méthodologie et les limites doivent être bien documentées au moins une fois par année. Toute amélioration ou modification doit être expliquée. L'information dérivée de concordances ou de conversions doit aussi être documentée dans les rapports. En ce qui concerne les bases de données dont sont extraites des données à tout moment, il est essentiel que la documentation sur les concordances et les conversions soit toujours à jour. Enfin, il n'est pas nécessaire de documenter les concordances et les conversions simples, comme celles des groupes d'âge d'un an aux groupes d'âge de cinq ans.

Ce critère est respecté si les concordances ou les conversions sont bien documentées au moins une fois par année.

**Critère 42** *L'ampleur des problèmes liés aux concordances et aux conversions se classe parmi l'une des catégories prédéterminées.*

Ce critère évalue les concordances et les conversions utilisées dans la base de données. En général, toute concordance ou conversion devrait être testée rigoureusement avant d'être appliquée à une base de données. Les erreurs de classification devraient être analysées et corrigées au besoin. Outre l'expérience concrète, il pourrait être utile de consulter la documentation pertinente sur les problèmes connus. Si la base de données utilise plus d'une concordance ou conversion, il faut réaliser l'évaluation globale à partir de celle dont la cote est la plus faible. Ce critère doit être évalué à partir des lignes directrices suivantes.

Cote proposée	Ligne directrice
Minimal	Peu ou pas de problèmes
Modéré	Problèmes identifiables à portée limitée
Significatif	Les données sont en grande partie mal converties et cela influe sur les résultats
Inconnu	L'équivalence n'a pas été examinée

## Comparabilité rétrospective

### Critères

- 43 *La documentation des modifications antérieures apportées à la banque de données existe et est facilement accessible.*
- 44 *Le réalisation d'une analyse des tendances permet d'examiner les changements dans les éléments de données de base au fil du temps.*
- 45 *L'ampleur des problèmes découlant de la comparaison des données au fil du temps se classe parmi l'une des catégories prédéterminées.*

On entend par comparabilité rétrospective l'uniformité des concepts et des méthodes de données dans le temps qui permet de faire des comparaisons valides de différentes estimations à différents moments dans le temps. Plusieurs facteurs peuvent compliquer la comparaison de données au fil du temps. Des modifications apportées à une base de données dans le but de l'améliorer peuvent parfois entraver la comparabilité rétrospective. Dans ces situations, cette limite devrait être décrite dans la documentation à l'intention des utilisateurs de données.

**Critère 43** *La documentation des modifications antérieures apportées à la banque de données existe et est facilement accessible.*

Ce critère évalue s'il existe une documentation des modifications antérieures et si elle est groupée en un document. Elle devrait inclure les changements apportés aux concepts, aux méthodologies, aux bases de sondage et aux éléments de données. Veuillez noter qu'un ensemble de manuels, chacun décrivant les changements de l'année en cours, n'est pas une forme acceptable de documentation rétrospective, car il devient trop difficile de suivre l'évolution des changements. Par ailleurs, le stockage de commentaires dans un programme SAS ne constitue pas non plus une forme acceptable de documentation s'il n'y a aucune autre documentation rétrospective. Les principaux changements des années précédentes devraient être inclus dans la documentation sur la qualité des données, mais un document plus détaillé pour usage interne peut aussi être nécessaire. Il est recommandé d'inclure de tels renseignements dans la documentation sur les métadonnées; veuillez vous rapporter à la section 4.3 pour obtenir plus de détails.

Ce critère est respecté si un seul document sur les modifications antérieures existe et qu'il est mis à jour chaque année.

**Critère 44** *La réalisation d'une analyse des tendances permet d'examiner les changements dans les éléments de données de base au fil du temps.*

L'analyse des tendances comprend des comparaisons de chiffres ou de pourcentages au fil du temps, de même que l'analyse de séries chronologiques plus complexes, du lissage ou de l'ajustement de courbes. La mise en graphique est souvent particulièrement utile pour étudier des variations temporelles. Dans une base de données sur les médecins, on pourrait examiner les fluctuations du nombre de médecins dans un secteur géographique particulier au cours des années précédentes. L'une des principales raisons d'être de l'analyse longitudinale est de déceler tout problème potentiel dans les données qui résulte de modifications apportées aux concepts ou aux méthodologies. Il est à noter que l'absence de changement au fil des ans peut aussi indiquer un problème si, selon les prévisions, les données devraient poursuivre une tendance naturelle à la hausse ou à la baisse en raison des politiques mises en œuvre ou des changements sociaux ou économiques.

Ce critère est respecté si une analyse des tendances a été réalisée pour les éléments de données de base depuis la dernière évaluation de la qualité des données.

**Critère 45** *L'ampleur des problèmes découlant de la comparaison des données au fil du temps se classe parmi l'une des catégories prédéterminées.*

Il est important de tenir compte des difficultés inhérentes à la production d'estimations tendanciennes valides. En cas de changements dans la méthodologie, les critères d'inclusion ou la non-réponse totale, il peut être impossible de déterminer si les changements observés étaient réels ou non. Par exemple, le nombre total d'admissions calculé pour un établissement de soins de courte durée peut être trompeur les fusions ou les modifications au type d'établissement ne sont pas prises en considération. Un changement dans les critères d'inclusion reposant sur le montant total facturé à la province peut invalider les estimations antérieures du nombre de médecins travaillant dans une province. Voici des lignes directrices générales pour l'évaluation de ce critère.

Cote proposée	Ligne directrice
Minimal	Peu ou pas de problèmes dans l'établissement de tendances comparables
Modéré	Problèmes trouvés dans quelques données tendanciennes
Significatif	Des données tendanciennes exactes n'ont pu être obtenues pour un élément de données de base
Inconnu	Il n'a pu être déterminé si des tendances exactes peuvent être établies

## 4 Dimension de la facilité d'utilisation

La facilité d'utilisation désigne la facilité avec laquelle on peut comprendre les données d'une banque de données et y accéder. Des données ou d'autres produits d'information difficiles à utiliser peuvent perdre toute valeur, quel que soit leur degré d'exactitude, d'actualité, de comparabilité ou de pertinence.

Plusieurs facteurs contribuent à la facilité d'utilisation des données d'une banque de données. En général, plus les limites et les exceptions associées aux données sont nombreuses, plus il est difficile d'interpréter les données. Les efforts accomplis pour améliorer la normalisation des données accroissent non seulement la facilité d'utilisation des données, mais aussi leur exactitude. L'application de méthodes non uniformes peut aussi compliquer l'interprétation des données. Les avantages découlant de l'instauration de nouvelles méthodes (p. ex. modifications du nom d'éléments de données ou de définitions) devraient donc être évalués en fonction des pertes d'interprétabilité. En d'autres mots, moins il y a de limites et de changements, plus il est facile d'interpréter les données.

Pour aider les principaux utilisateurs à interpréter les données, il faut les informer de toute limite majeure connue au moment de la diffusion des données et sur une base régulière après la diffusion. Une fois les principales limites connues, elles devraient être documentées pour les utilisateurs. Il en est de même pour les méthodes des banques de données et les changements aux méthodes. Les données doivent aussi être présentées dans un format facilement accessible et convivial. Enfin, même si une banque de données est bien documentée et accessible, ses données ne serviront pas si les utilisateurs n'en connaissent pas l'existence.

La dimension de la facilité d'utilisation vise à cerner les aspects problématiques d'une banque de données sous l'angle de l'interprétabilité des données, de même qu'à déterminer si les données sont bien documentées et accessibles. Elle comporte les caractéristiques suivantes :

- Accessibilité (les données sont-elles facilement accessibles?)
- Documentation (les données sont-elles bien documentées?)
- Interprétabilité (les données sont-elles faciles à comprendre?)

Dimension	Caractéristiques	Critères
Facilité d'utilisation	Accessibilité	46 à 48
	Documentation	49 à 51
	Interprétabilité	52 et 53

## Accessibilité

### Critères

46 *Un ensemble de données final est rendu disponible à la date prévue de diffusion.*

47 *Des tableaux et des analyses au format et au contenu standard sont produits lors de chaque diffusion prévue ou sur demande.*

48 *Les produits sont définis, catalogués ou annoncés.*

Cette caractéristique porte sur la facilité avec laquelle les données d'une banque de données peuvent être obtenues de l'ICIS. Elle comprend la facilité avec laquelle on peut déterminer l'existence de la banque de données ainsi que l'adéquation du format des données. Les données que les utilisateurs ne connaissent pas, ne peuvent trouver ou ne peuvent importer dans leur propre environnement de travail pour quelque raison que ce soit ne leur seront d'aucune utilité.

**Critère 46** *Un ensemble de données final est rendu disponible à la date prévue de diffusion.*

Les données utilisées dans les analyses et la création de rapports devraient être sauvegardées dans un endroit sécuritaire pour consultation future. Il faut souvent se référer à des ensembles de données antérieurs pour réaliser d'autres analyses. La sauvegarde d'une version de l'ensemble de données utilisé pour la création d'un rapport permet de s'assurer que les résultats découlant d'une nouvelle analyse sont cohérents avec les résultats diffusés antérieurement. Veuillez noter que les données peuvent être fournies sous des formats différents, selon les demandes des utilisateurs. Lorsque des analyses sont effectuées directement sur une base de données, il est important de conserver les codes SAS utilisés pour produire les tableaux pendant un an suivant la diffusion du rapport.

Ce critère est respecté si le fichier de données d'une banque de données (ou le code SAS s'il est impossible de produire un ensemble de données) est mis à la disposition des intervenants à la suite d'une diffusion prévue.

**Critère 47** *Des tableaux et des analyses au format et au contenu standard sont produits lors de chaque diffusion prévue ou sur demande.*

Pour beaucoup d'utilisateurs, des statistiques agrégées ou des tableaux récapitulatifs sont plus utiles que des microdonnées. Outre les rapports principaux, les utilisateurs devraient aussi avoir accès aux statistiques agrégées et aux tableaux standard pour chaque diffusion prévue. Les tableaux standard sont habituellement des tableaux croisés des éléments de données de base de la base de données. Il faut comparer les résultats tirés des tableaux standard à ceux des années précédentes ou les diffuser pour confirmer leur vraisemblance. Pour certains utilisateurs, le fichier de microdonnées peut contenir l'information la plus utile. Dans de telles situations, il est important de respecter le plus possible les formats précédents afin de permettre les comparaisons.

Ce critère est respecté si les analyses et les tableaux standard généralement utilisés sont disponibles sur demande ou pour chaque diffusion prévue.

**Critère 48** *Les produits sont définis, catalogués ou annoncés.*

Pour aider les utilisateurs, il faut inscrire les produits d'une banque de données (rapports annuels, rapports analytiques, rapports personnalisés, etc.) dans les systèmes de diffusion de l'organisme, comme le site Web de l'ICIS, le calendrier des produits livrables de l'ICIS et le Guide des produits et services de l'ICIS. Le site Web peut servir de bibliothèque virtuelle pour tous les produits d'information de l'ICIS accessibles au public. Sur le site Web, il est possible de faire une demande par l'intermédiaire du Programme d'accès aux données pour les étudiants des cycles supérieurs (PADECS) et du Service des applications spéciales (AS), ou encore de faire à partir des options sur les demandes spéciales de recherche et de données brutes. Il est aussi possible d'avoir recours à d'autres canaux, comme les médias (communiqués), les centres de données de recherche et les bibliothèques publiques, pour recueillir de l'information sur les banques de données de l'ICIS.

Ce critère est respecté si les produits d'une banque de données sont définis, catalogués ou annoncés dans n'importe lequel des systèmes de diffusion de l'ICIS pour chaque diffusion prévue.

## Documentation

### Critères

*49 Il existe de la documentation sur la qualité des données à l'intention des utilisateurs.*

*50 Il existe de la documentation sur les métadonnées.*

*51 Un avertissement accompagne toute diffusion préliminaire.*

Cette caractéristique permet de déterminer si la documentation requise pour comprendre les données est disponible. La documentation est nécessaire pour l'interprétation et l'utilisation appropriées des données d'une banque de données. Cette documentation inclut notamment des renseignements sur le contexte, les sources de données et la méthodologie, les concepts et les éléments de données mesurés, ainsi que sur l'exactitude et la comparabilité des données (voir la section 4 pour de plus amples renseignements). Tout au long du cycle de travail de la qualité des données (mentionné à la section 2), il faut s'assurer de documenter régulièrement les différentes étapes de planification, de mise en œuvre et d'évaluation et de documenter clairement pour les utilisateurs l'objet, la méthode et la qualité de la mesure.

**Critère 49** *Il existe de la documentation sur la qualité des données à l'intention des utilisateurs.*

La documentation sur la qualité des données à l'intention des utilisateurs a pour but de donner aux utilisateurs internes et externes de la banque de données suffisamment d'information pour qu'ils puissent décider si la qualité des données convient à ce qu'ils désirent en faire. Toute diffusion devrait également inclure les coordonnées des personnes-ressources afin que les utilisateurs puissent obtenir l'information supplémentaire dont ils ont besoin sur les limites, selon l'usage qu'ils désirent faire des données.

Un document indépendant sur la qualité des données devrait être mis à la disposition des utilisateurs au moins une fois par année. Veuillez consulter la section 4.2 pour obtenir de plus amples renseignements au sujet de la documentation sur la qualité des données à l'intention des utilisateurs.

Ce critère est respecté si une documentation à jour sur la qualité des données est disponible chaque fois que des données sont diffusées ou extraites.

NOUVEAU

**Critère 50** *Il existe de la documentation sur les métadonnées.*

Pour faciliter l'interprétation et l'utilisation appropriée des données, il faut mettre à la disposition du personnel de l'ICIS qui travaille avec les données toute documentation sur les méthodes des banques de données. À la section 4.3, ce type de documentation est appelé « documentation sur les métadonnées ». Même si la documentation sur la qualité des données présente des renseignements généraux et souligne les limites principales des données, la documentation interne sur les métadonnées doit contenir des notes contextuelles plus détaillées, de même que toutes les limites connues à l'égard des données.

Ce critère est respecté s'il existe une documentation sur les métadonnées de la banque de données aux fins d'utilisation interne. Cette documentation devrait être révisée sur une base annuelle et mise à jour au besoin.

**Critère 51** *Un avertissement accompagne toute diffusion préliminaire.*

Afin d'améliorer l'actualité, certaines banques de données peuvent diffuser des données ou des résultats préliminaires. Si la banque de données ne diffuse pas de données préliminaires ou de rapports basés sur des données préliminaires, ce critère ne s'applique pas.

Une **diffusion préliminaire** peut inclure la diffusion de données conçues pour authentifier ou valider des données. Par exemple, avant leur publication, des chiffres sur les indicateurs de santé pourraient être envoyés aux régions sanitaires aux fins de vérification et de validation. On entend aussi par **diffusion préliminaire** une diffusion peut-être incomplète de données visant à raccourcir les délais de publication. Par exemple, des données recueillies annuellement pourraient être diffusées six mois avant la fin de l'année afin que les planificateurs des systèmes de soins de santé aient un premier aperçu des données complètes qui suivront.

Les diffusions préliminaires doivent être accompagnées d'un avertissement précisant que les données ne sont peut-être pas complètes et qu'elles sont sujettes à révision. Une description des taux de réponse totale et partielle, de même que des taux de réponse totale et partielle finaux prévus, devrait également être incluse. Il faut également préciser les révisions prévues et leur incidence possible.

Ce critère est respecté si un avertissement accompagne toute diffusion préliminaire de données.

## Interprétabilité

### Critères

*52 Un mécanisme permet aux principaux utilisateurs de communiquer leurs commentaires à la section d'une banque de données et de recevoir une réponse de cette dernière.*

*53 Des lignes directrices de révision sont disponibles et appliquées à chaque diffusion.*

On entend par interprétabilité la facilité avec laquelle l'utilisateur peut comprendre les données. L'interprétabilité dépend en grande partie des caractéristiques du plan et des limites sous-jacentes de la qualité des données. Par exemple, une population de référence complexe limite non seulement la généralisation des données, mais peut aussi limiter la facilité avec laquelle les données peuvent être comprises. Si des concepts et des classifications standard sont en place, il sera plus facile de comprendre et d'utiliser les données. Pour faciliter l'interprétabilité, il peut être utile de joindre la disposition des enregistrements aux différents fichiers de données.

Comme il est difficile de mesurer directement le concept d'interprétabilité, cette caractéristique mesure si des lignes directrices de révision et un mécanisme ont été mis en place en vue de faciliter l'interprétation.

**Critère 52** *Un mécanisme permet aux principaux utilisateurs de communiquer leurs commentaires à la section d'une banque de données et de recevoir une réponse de cette dernière.*

Les coordonnées (nom, numéro de téléphone et courriel) des personnes-ressources doivent être incluses dans chaque diffusion afin que les utilisateurs (internes ou externes) puissent faire part de leur rétroaction concernant les principales limites de la qualité des données au moment où elles sont connues. Les principaux utilisateurs devraient être invités à informer les personnes-ressources des limites qu'ils découvrent et de toute préoccupation.

De la même façon, il devrait y avoir un mécanisme pour informer les principaux utilisateurs des limites découvertes après la diffusion, des mesures prises à cet égard et de l'incidence des erreurs. Des exemples d'un tel mécanisme pourraient comprendre un système d'avis pour les utilisateurs ou un groupe formé d'utilisateurs principaux (p. ex. Statistique Canada, Santé Canada et des analystes de l'ICIS).

Ce critère est respecté si les coordonnées des personnes-ressources de la section de la banque de données sont incluses dans toute diffusion importante de données ou de rapports et que les utilisateurs sont invités à donner leur rétroaction.

**Critère 53** *Des lignes directrices de révision sont disponibles et appliquées à chaque diffusion.*

Il arrive que les estimations initiales soient révisées à mesure que des erreurs ou des données manquantes sont découvertes. Une révision peut s'avérer nécessaire lorsque les erreurs ou les données manquantes sont corrigées. Une **révision** est une modification apportée aux données, ou aux estimations tirées des données, une fois que celles-ci sont du domaine public. Si des limites ou des mises à jour importantes sont découvertes après la diffusion, des lignes directrices propres à chaque base de données devraient permettre de décider s'il faut ou non diffuser les données révisées. Les lignes directrices devraient indiquer plus particulièrement à quel moment l'incidence d'erreurs nouvellement découvertes ou de mises à jour serait suffisamment importante pour justifier la diffusion d'un sous-ensemble de données révisées. Ces lignes directrices devraient aussi indiquer comment et quand les révisions sont normalement publiées. Par exemple, une ligne directrice de révision pourrait énoncer que lorsque des estimations nationales sont fortement touchées par une correction après diffusion, un ensemble de données révisées est diffusé.

Ce critère est respecté si des lignes directrices de révision propres à une banque de données sont disponibles et appliquées chaque fois que des données sont diffusées ou extraites. Ces lignes directrices doivent être à jour et présentées sous forme de document.

## 5 Dimension de la pertinence

La pertinence décrit de quelle façon une banque de données répond aux besoins actuels et potentiels des utilisateurs. Pour assurer la pertinence, il faut rester en contact avec les principaux utilisateurs et intervenants. La pertinence évalue si les données disponibles traitent des questions les plus importantes pour les utilisateurs. Outre son souci de fournir des données exactes, opportunes, comparables et utilisables, l'ICIS, pour remplir son mandat, doit aussi faire en sorte que ses banques de données reflètent continuellement les besoins les plus importants en matière d'information sur les soins de santé au Canada. La difficulté consiste à atteindre un point d'équilibre entre les différents besoins des utilisateurs actuels et potentiels en vue de produire un programme qui satisfait le plus possible les principaux besoins.

La dimension de la pertinence vise à évaluer dans quelle mesure une banque de données s'adapte au changement et si elle est perçue comme étant utile. La pertinence comporte les caractéristiques suivantes :

- Adaptabilité (peut-on prévoir les besoins des utilisateurs et planifier en conséquence?)
- Valeur (quelle valeur ont les données?)

Dimension	Caractéristiques	Critères
Pertinence	Adaptabilité	54 et 55
	Valeur	56 à 58

### Adaptabilité

#### Critères

*54 Des mécanismes tiennent les clients et les intervenants informés des développements dans le domaine.*

*55 La banque de données est conçue de manière à faciliter les modifications futures du système.*

L'adaptabilité désigne la situation d'une banque de données et sa souplesse à répondre aux besoins en information actuels et futurs de ses principaux utilisateurs. Pour demeurer pertinente, une banque de données devra peut-être s'adapter continuellement aux nouvelles questions dans le domaine. Comme les besoins et les priorités changent sans cesse, des mécanismes de rétroaction permettraient de sensibiliser les principaux clients et groupes d'intervenants aux questions d'intérêt actuelles et futures.

Lorsque les questions actuelles ou émergentes sont connues et surveillées, il est possible de prévoir les besoins en information. Il est important de demeurer proactif et non pas réactif face aux besoins des principaux utilisateurs. La conception de la banque de données peut tenir compte des besoins anticipés en information. Même s'il est impossible de prévoir les besoins des utilisateurs avec une exactitude absolue, il est utile de tenter de concevoir des banques de données capables de répondre au changement. Il est également important que la banque de données ne soit pas trop souple par rapport aux volontés et aux besoins des utilisateurs. L'incidence des changements sur l'ensemble du cycle du travail devrait faire l'objet d'une évaluation complète avant leur intégration.

**Critère 54** *Des mécanismes tiennent les clients et les intervenants informés des développements dans le domaine.*

Le maintien d'un lien avec les intervenants par le personnel de la section de la banque de données permet à l'ICIS de se tenir au courant des questions actuelles et émergentes et des besoins en information susceptibles d'en découler. Le personnel de la section peut demeurer en contact avec les principaux clients ou intervenants en organisant des réunions de groupes d'experts, de comités directeurs et de comités consultatifs professionnels, ou en y participant. La participation à des conférences et la soumission des documents à des revues évaluées par des pairs peuvent aussi être instructives. Les systèmes d'interrogation en ligne (le cas échéant) peuvent également servir de mécanisme pour tenir informés les clients ou les intervenants. Par exemple, l'outil d'interrogation en ligne de l'ICIS répond aux questions des utilisateurs au sujet de la codification. Si les banques de données intéressent les mêmes utilisateurs et intervenants, il est possible d'envisager une approche coordonnée du processus de consultation pour toutes les banques de données.

Ce critère est respecté si des mécanismes de liaison permettent aux clients et aux intervenants de demeurer à la fine pointe des développements dans le domaine.

**Critère 55** *La banque de données est conçue de manière à faciliter les modifications futures du système.*

Pour qu'une banque de données demeure pertinente, des modifications continues peuvent s'avérer nécessaires. De plus, pour répondre aux questions émergentes, il peut être nécessaire de redéfinir des éléments de données existants ou d'en ajouter de nouveaux. Par exemple, il est possible d'ajouter de nouveaux éléments de données sur la date et l'heure pour recueillir des données sur les temps d'attente dans les services d'urgence dans une base de données sur les soins ambulatoires. Une banque de données devrait aussi pouvoir intégrer de nouvelles normes techniques (p. ex. la CIM-10-CA).

Des modifications aux banques de données peuvent s'avérer nécessaires non seulement pour se pencher sur les importantes questions émergentes ou les nouvelles normes techniques, mais aussi pour faire face aux limites de la qualité des données. Par exemple, si des durées de séjour négatives sont observées, il est possible d'ajouter de nouvelles vérifications. L'adaptation aux questions émergentes, l'intégration de nouvelles normes et l'examen des questions de qualité des données sous le contrôle de la section d'une banque de données pourraient bien être considérés comme faisant partie du processus permanent d'amélioration d'une banque de données.

Bien que la souplesse d'une banque de données soit importante, les avantages de toute modification devraient être évalués en fonction des pertes potentielles de comparabilité ou d'interprétabilité.

Ce critère est respecté si la banque de données a démontré sa capacité de s'adapter à une importante question émergente, à une nouvelle norme technique ou à une limite marquée de la qualité des données.

## Valeur

### Critères

*56 Le mandat de la banque de données comble une lacune en matière d'information sur la santé.*

*57 Le niveau d'utilisation de la banque de données est surveillé.*

*58 La satisfaction des utilisateurs est évaluée périodiquement.*

On définit la valeur d'une banque de données par sa contribution aux connaissances sur la santé de la population ou le système de santé et par son utilisation. Autrement dit, la valeur ou l'utilité d'une banque de données dépend des deux aspects suivants : La banque de données comble-t-elle un manque d'information sur la santé ou sur le système de santé? Remplit-elle son mandat?

Outre la communication continue avec les principaux utilisateurs et intervenants pour mieux connaître les besoins émergents en information, il faudrait aussi surveiller la valeur perçue accordée aux données d'une banque de données. Les mécanismes de liaison décrits précédemment (critère 54) devraient être utilisés pour obtenir de la rétroaction sur les programmes actuels et sur les besoins à venir.

**Critère 56** *Le mandat de la banque de données comble une lacune en matière d'information sur la santé.*

La valeur d'une banque de données dépend de sa capacité à combler une lacune en information sur la santé. Le mandat d'une banque de données devrait être évalué de façon périodique en fonction des autres banques de données de l'ICIS et dans tout le secteur externe. Il importe de bien comprendre comment une banque de données complète les autres banques de données de l'ICIS et comment elle se compare à d'autres sources de données similaires dans le domaine.

Ce critère est respecté si la banque de données comble une lacune en information dans le secteur des soins de santé.

**Critère 57** *Le niveau d'utilisation de la banque de données est surveillé.*

La valeur d'une banque de données peut être liée à l'étendue de l'utilisation des données. Des preuves de l'utilisation pourraient inclure : des utilisations de données qui retiennent l'attention du public (p. ex. le rapport Romanow), des visites de pages Web, des coupures de presse, des annonces, des citations, des ouvrages rédigés par le personnel, des ventes, des communications avec le personnel et des apparitions du personnel dans les médias, dans des conférences et dans des forums stratégiques. Le niveau d'utilisation peut aussi être vérifié grâce au suivi du nombre et du type de demandes de données et, dans la mesure du possible, grâce à des rapports électroniques.

Ce critère est respecté si le niveau d'utilisation des banques de données est vérifié.

**Critère 58** *La satisfaction des utilisateurs est évaluée périodiquement.*

Il est important d'évaluer si la banque de données répond aux besoins des utilisateurs et d'intégrer les résultats de l'évaluation dans une révision du programme. Les résultats de sondages sur la satisfaction des intervenants peuvent servir de preuve directe de la valeur perçue de la banque de données. Un sondage sur la satisfaction des clients peut aussi être un bon moment pour leur demander ce qu'ils pensent de l'exactitude, de l'actualité, de la comparabilité et de la facilité d'utilisation des données. Les analystes internes sont une source importante de rétroaction; c'est pourquoi les résultats de toutes les analyses internes pertinentes devraient être examinés attentivement, y compris les révisions de la foire aux questions (FAQ) si un outil de requête ou une ligne de dépannage sont offerts aux clients. Un sondage sur la satisfaction des clients peut aussi révéler qu'une base de données dont la déclaration de données est obligatoire possède une valeur beaucoup plus grande qu'une base de données reposant sur la soumission de données sur une base volontaire. En raison de contraintes liées aux ressources, une telle révélation peut alors mener à la décision de sélectionner un échantillon au sein d'une population de référence et de s'attendre à ce que cet échantillon soumette ses données sur une base obligatoire. Dans tous les cas, l'incidence des modifications suggérées dans un sondage sur la satisfaction des intervenants doit faire l'objet d'une évaluation complète avant toute mise en œuvre.

Ce critère est respecté si des évaluations périodiques de la satisfaction des clients sont réalisées.



**NOUVEAU**

## **Annexe D – Modèle de rapport d'évaluation de la qualité des données**

**ICIS**

**Rapport d'évaluation de la qualité  
des données pour  
[Base de données]  
[année des données]**

**[Année Mois Jour]**

d'après le cadre de la qualité des données de l'ICIS 2009

## Sommaire

Un sommaire doit être inclus pour résumer les points forts de la banque de données et souligner les secteurs problématiques. Il est également possible de présenter les résultats sous forme de tableau, comme dans l'exemple ci-dessous.

### Résumé des critères d'évaluation

Dimensions et caractéristiques	Critères	Évaluation
<b>Exactitude</b>		
<b>Couverture</b>	1a La population de référence est explicitement mentionnée dans toutes les diffusions.	
	1b Des efforts sont faits pour réduire l'écart entre la population de référence et la population cible.	
	2 Les sources connues de sous-dénombrement ou de surdénombrement ont été documentées.	
	3 La base de sondage a été validée au moyen d'une comparaison des sources externes et indépendantes.	
	4 Le taux de sous-dénombrement ou de surdénombrement se classe parmi l'une des catégories prédéfinies.	
<b>Saisie et collecte</b>	5a Les pratiques de l'ICIS qui minimisent le fardeau de réponse sont documentées.	
	5b L'ICIS possède de la documentation sur les pratiques des fournisseurs de données qui minimisent le fardeau de réponse.	
	6 Il existe des pratiques qui favorisent la collaboration pour la soumission de données.	
	7 Il existe des pratiques qui appuient les fournisseurs de données.	
	8 Il existe des processus de soumission standard et ils sont suivis par les fournisseurs de données.	
	9 Il existe des mesures de contrôle de la qualité de la saisie des données et elles sont mises en œuvre par les fournisseurs de données.	
<b>Non-réponse totale</b>	10 L'ampleur de la non-réponse totale est mentionnée dans la documentation sur la qualité des données.	
	11 Le nombre d'enregistrements reçus pour les unités déclarantes est examiné afin que les valeurs inhabituelles puissent être décelées.	
	12 L'ampleur de la non-réponse totale se classe parmi l'une des catégories prédéterminées.	
<b>Non-réponse partielle (ou non-réponse par question)</b>	13 La non-réponse partielle est connue.	
	14 L'ampleur de la non-réponse partielle se classe parmi l'une des catégories prédéterminées.	

Dimensions et caractéristiques	Critères	Évaluation
<b>Erreur de mesure</b>	15 Le niveau de l'erreur de mesure se classe parmi l'une des catégories prédéterminées.	
	16 Le niveau du biais n'est pas significatif.	
	17 Le degré de problèmes de cohérence se classe parmi l'une des catégories prédéterminées.	
<b>Vérification et imputation</b>	18 Des contrôles de la validité sont réalisés pour chacun des éléments de données et toutes les données invalides sont signalées.	
	19 Les règles de vérification et l'imputation sont logiques et appliquées de façon uniforme.	
	20 Les rapports de vérification destinés aux utilisateurs sont faciles à utiliser et à comprendre.	
	21 Le processus d'imputation est automatisé et conforme aux règles de vérification.	
<b>Traitement et estimation</b>	22 La documentation sur toutes les activités de traitement des données est tenue à jour.	
	23 Les spécifications techniques pour la banque de données sont tenues à jour.	
	24 Les modifications à la structure sous-jacente ou aux programmes de traitement ou d'estimation d'une banque de données ont été testées.	
	25 Les données brutes, conformément à la politique de l'ICIS en matière de maintien des données, sont sauvegardées dans un endroit sécuritaire.	
	26a Les statistiques agrégées d'une banque de données ont été comparées, dans la mesure du possible, à des statistiques similaires provenant d'une autre banque de données de l'ICIS ou d'une source externe.	
	26b La variance de l'estimation, comparée à l'estimation elle-même, se situe à un niveau acceptable.	

<b>Dimensions et caractéristiques</b>	<b>Critères</b>	<b>Évaluation</b>
<b>Actualité</b>		
<b>Actualité des données au moment de leur diffusion</b>	27 L'écart entre la date réelle de diffusion et la fin de la période de référence est raisonnablement bref.	
	28 La date officielle de diffusion a été annoncée avant la diffusion.	
	29 La diffusion a eu lieu à la date officielle prévue.	
	30 Les activités de traitement des données sont revues régulièrement aux fins d'amélioration de l'actualité des données.	
<b>Actualité de la documentation</b>	31 La documentation recommandée sur la qualité des données était disponible au moment de la diffusion des données ou des rapports.	
	32 Les principaux rapports ont été diffusés à la date prévue.	
<b>Comparabilité</b>		
<b>Normes du Dictionnaire de données</b>	33 Tous les éléments de données font l'objet d'une évaluation visant à déterminer s'ils sont inclus dans le Dictionnaire de données de l'ICIS.	
	34 Les éléments de données d'une banque de données qui sont inclus dans le Dictionnaire de données de l'ICIS doivent être conformes aux normes de ce dictionnaire.	
<b>Normalisation</b>	35 Les données sont recueillies au niveau de détail le plus élevé possible.	
	36 Pour tout élément de données dérivé, l'élément de données original demeure accessible.	
<b>Couplage</b>	37 Les données géographiques sont recueillies au moyen de la Classification géographique type (CGT).	
	38 Les données sont recueillies selon le même calendrier, plus particulièrement d'une province ou d'un territoire à l'autre et au sein d'une province ou d'un territoire.	
	39 L'utilisation d'identificateurs permet de distinguer les établissements ou les organismes seulement aux fins de couplage rétrospectif.	
	40 L'utilisation d'identificateurs permet de distinguer les personnes ou les appareils seulement aux fins de couplage rétrospectif.	
<b>Équivalence</b>	41 La méthodologie et les limites des concordances ou des conversions sont documentées.	
	42 L'ampleur des problèmes liés aux concordances et aux conversions se classe parmi l'une des catégories prédéterminées.	

<b>Dimensions et caractéristiques</b>	<b>Critères</b>	<b>Évaluation</b>
<b>Comparabilité rétrospective</b>	43 La documentation des modifications antérieures apportées à la banque de données existe et est facilement accessible.	
	44 La réalisation d'une analyse des tendances permet d'examiner les changements dans les éléments de données de base au fil du temps.	
	45 L'ampleur des problèmes découlant de la comparaison des données au fil du temps se classe parmi l'une des catégories prédéterminées.	
<b>Facilité d'utilisation</b>		
<b>Accessibilité</b>	46 Un ensemble de données final est rendu disponible à la date prévue de diffusion.	
	47 Des tableaux et des analyses au format et au contenu standard sont produits lors de chaque diffusion prévue ou sur demande.	
	48 Les produits sont définis, catalogués ou annoncés.	
<b>Documentation</b>	49 Il existe de la documentation sur la qualité des données à l'intention des utilisateurs.	
	50 Il existe de la documentation sur les métadonnées.	
	51 Un avertissement accompagne toute diffusion préliminaire.	
<b>Interprétabilité</b>	52 Un mécanisme permet aux principaux utilisateurs de communiquer leurs commentaires à la section d'une banque de données et de recevoir une réponse de cette dernière.	
	53 Des lignes directrices de révision sont disponibles et appliquées à chaque diffusion.	
<b>Pertinence</b>		
<b>Adaptabilité</b>	54 Des mécanismes tiennent les clients et les intervenants informés des développements dans le domaine.	
	55 La banque de données est conçue de manière à faciliter les modifications futures du système.	
<b>Valeur</b>	56 Le mandat de la banque de données comble une lacune en matière d'information sur la santé.	
	57 Le niveau d'utilisation de la banque de données est surveillé.	
	58 La satisfaction des utilisateurs est évaluée périodiquement.	

## Recommandations

Les recommandations formulées tout au long du rapport doivent être résumées dans un plan d'action figurant dans le sommaire. Le plan d'action indique les recommandations, la date de mise en œuvre, la date butoir et les responsables. De plus, il précise si la participation de la Division de la qualité des données est requise. Vous trouverez ci-dessous un modèle de plan d'action à remplir.

### Plan d'action

NOUVEAU

Recommandation	Date de mise en œuvre	Responsables	Date butoir ou processus continu	Participation de la Division de la qualité des données
1.				
2.				
3.				
4.				
5.				

## Introduction

Il est recommandé de commencer le rapport par une brève introduction décrivant la banque de données et le délai d'exécution de l'évaluation.

## Évaluation détaillée

Les dimensions, les caractéristiques et les critères devraient être indiqués dans la présente section. Les cotes pour chaque critère doivent être appuyées par au moins une ou deux phrases décrivant pourquoi la cote a été attribuée. De plus, lorsque la situation le permet, des documents à l'appui doivent être inclus ou, du moins, donnés en référence.

# 1 Dimension de l'exactitude

## 1.1 Couverture

**Critère 1a** *La population de référence est explicitement mentionnée dans toutes les diffusions.*

Évaluez ce critère ici.

**Détails requis** : donnez d'abord des détails sur les aspects suivants de la banque de données : population cible, population de référence, unités de la base de sondage, unités d'analyse et période de référence. Ensuite, indiquez si la population de référence est explicitement mentionnée dans toutes les diffusions. Indiquez aussi le titre des documents diffusés et l'endroit à partir duquel ils peuvent être consultés. Si la population de référence mentionnée dans les diffusions varie d'une année à l'autre ou d'un document à l'autre, donnez des détails à ce sujet.

**Évaluation** :  respecté  non respecté

**Critère 1b** *Des efforts sont faits pour réduire l'écart entre la population de référence et la population cible.*

Évaluez ce critère ici.

**Détails requis** : si une différence entre la population de référence et la population cible a été soulevée, expliquez quels efforts ont été déployés pour réduire l'écart entre ces deux populations.

**Évaluation** :  respecté  non respecté  inconnu  non applicable

**Critère 2** *Les sources connues de sous-dénombrement ou de surdénombrement ont été documentées.*

Évaluez ce critère ici.

**Détails requis** : donnez les sources et les raisons expliquant le sous-dénombrement et le surdénombrement. Veuillez noter que tous les niveaux d'observation (c'est-à-dire les unités de la base de sondage) doivent être indiqués comme le décrit le tableau B de la section Dimension de l'exactitude (annexe C).

**Évaluation** :  respecté  non respecté  inconnu

**Critère 3** *La base de sondage a été validée au moyen d'une comparaison des sources externes et indépendantes.*

Évaluez ce critère ici.

**Détails requis** : expliquez les efforts faits pour valider la base de sondage au moyen de comparaisons entre des sources externes indépendantes et la banque de données. Donnez des détails sur les sources utilisées et leur crédibilité. Veuillez noter qu'il est acceptable d'utiliser de multiples sources et d'effectuer une comparaison à un niveau agrégé.

**Évaluation** :  respecté  non respecté  non applicable

**Critère 4** *Le taux de sous-dénombrement ou de surdénombrement se classe parmi l'une des catégories prédéfinies.*

Évaluez ce critère ici.

**Détails requis** : décrivez les calculs utilisés pour déterminer le taux de surdénombrement ou de sous-dénombrement de la banque de données et donnez toute définition connexe. Veuillez noter que l'évaluation de ce critère dépend du résultat des calculs.

**Évaluation** :  aucun ou minimal (moins de 1 %)  
 modéré (de 1 à 5 %)  
 significatif (plus de 5 %)  
 inconnu (le résultat n'a pu être déterminé)

## 1.2 Saisie et collecte

**Critère 5a** *Les pratiques de l'ICIS qui minimisent le fardeau de réponse sont documentées.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur les pratiques qu'emploie l'ICIS pour minimiser le fardeau de réponse, de même que le nom de la documentation sur ces pratiques et l'endroit où elle se trouve.

**Évaluation** :  respecté  non respecté  inconnu

**Critère 5b** *L'ICIS possède de la documentation sur les pratiques des fournisseurs de données qui minimisent le fardeau de réponse.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur les pratiques suivies par les fournisseurs de données pour minimiser le fardeau de réponse, de même que le nom de la documentation sur ces pratiques et l'endroit où elle se trouve.

**Évaluation** :  respecté  non respecté  inconnu

**Critère 6** *Il existe des pratiques qui favorisent la collaboration pour la soumission de données.*

Évaluez ce critère ici.

**Détails requis** : précisez la nature de la participation des fournisseurs de données (c'est-à-dire obligatoire ou volontaire) et les pratiques employées par l'ICIS pour favoriser la collaboration.

**Évaluation** :  respecté  non respecté  inconnu

**Critère 7** *Il existe des pratiques qui appuient les fournisseurs de données.*

Évaluez ce critère ici.

**Détails requis** : donnez des détails sur les pratiques qui appuient les fournisseurs de données avant et pendant la saisie des données (p. ex. les séances d'information, les réponses rapides aux courriels et aux appels téléphoniques, le soutien technique et l'aide à la codification, ainsi que l'accès à des documents d'aide, aux lignes directrices de codification et au manuel de saisie des données).

**Évaluation** :  respecté  non respecté  inconnu

**Critère 8** *Il existe des processus de soumission standard et ils sont suivis par les fournisseurs de données.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements au sujet des processus et des formulaires standard de soumission des données pour la banque de données, et précisez dans quelle mesure ils sont utilisés par les fournisseurs de données.

**Évaluation** :  respecté  non respecté  inconnu

**Critère 9** *Il existe des mesures de contrôle de la qualité de la saisie des données et elles sont mises en œuvre par les fournisseurs de données.*

Évaluez ce critère ici.

**Détails requis** : indiquez les mesures prises pour s'assurer que les données sont enregistrées adéquatement. Précisez si ces mesures sont mises en œuvre par le personnel de la banque de données ou les fournisseurs de données. Il peut être souhaitable d'indiquer si les fournisseurs de données sont capables d'appliquer de telles mesures et de mentionner les travaux réalisés par le personnel de la banque de données pour améliorer la saisie des données à l'échelle des fournisseurs.

**Évaluation** :  respecté  non respecté  inconnu

### 1.3 Non-réponse totale

**Critère 10** *L'ampleur de la non-réponse totale est mentionnée dans la documentation sur la qualité des données.*

Évaluez ce critère ici.

**Détails requis** : indiquez si l'ampleur de la non-réponse totale de la base de sondage est mentionnée dans la documentation sur la qualité des données fournie aux utilisateurs. (L'unité de la base de sondage peut être un patient, un dispensateur de services, une région sanitaire, une province ou tout autre fournisseur de données). Vous devez également préciser si la documentation est accessible, tant pour les utilisateurs que pour le personnel à l'interne, et comment y accéder.

**Évaluation** :  respecté  non respecté

**Critère 11** *Le nombre d'enregistrements reçus pour les unités déclarantes est examiné afin que les valeurs inhabituelles puissent être décelées.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur les procédures que suit la section de l'ICIS pour faire le suivi du nombre d'unités déclarantes au fil du temps. Indiquez les problèmes que posent ces procédures et les efforts déployés pour les régler.

**Évaluation** :  respecté  non respecté

**Critère 12** *L'ampleur de la non-réponse totale se classe parmi l'une des catégories prédéterminées.*

Évaluez ce critère ici.

**Détails requis** : décrivez les calculs utilisés pour déterminer le taux de non-réponse totale de la banque de données et donnez toute définition connexe. Comme ce critère tient compte tant du niveau de la base de sondage que du niveau de l'unité d'analyse, il est recommandé de calculer la non-réponse pour chaque niveau d'observation. Le personnel de l'ICIS calcule habituellement le taux de non-réponse totale plutôt que son complément, le taux de réponse totale. Veuillez noter que l'évaluation de ce critère dépend du résultat des calculs.

**Évaluation** :  aucun ou minimal (taux de non-réponse inférieur à 2 %)  
 modéré (de 2 à 10 %)  
 significatif (plus de 10 %)  
 inconnu (le résultat n'a pu être déterminé)

## 1.4 Non-réponse partielle (ou non-réponse par question)

**Critère 13** *La non-réponse partielle est connue.*

Évaluez ce critère ici.

**Détails requis** : nommez les processus et les programmes utilisés par la banque de données pour distinguer les champs vides des non-réponses aux éléments de données de base. Pour respecter ce critère, vous devez préciser le nom et l'emplacement du document dans lequel la non-réponse partielle est précisée.

**Évaluation** :  respecté  non respecté

**Critère 14** *L'ampleur de la non-réponse partielle se classe parmi l'une des catégories prédéterminées.*

Évaluez ce critère ici.

**Détails requis** : décrivez les calculs utilisés pour déterminer le niveau de non-réponse pour chaque élément de données de base et donnez toute définition connexe. Le personnel de l'ICIS calcule habituellement le taux de non-réponse partielle plutôt que son complément, le taux de réponse partielle. Lorsque vous évaluez ce critère, vous devez tenir compte de l'élément de données de base qui affiche le taux de non-réponse partielle le plus élevé.

**Évaluation** :  aucun ou minimal (taux de non-réponse partielle inférieur à 2 %)  
 modéré (de 2 à 10 %)  
 significatif (plus de 10 %)  
 inconnu (le résultat n'a pu être déterminé)

## 1.5 Erreur de mesure

**Critère 15** *Le niveau de l'erreur de mesure se classe parmi l'une des catégories prédéterminées.*

Évaluez ce critère ici.

**Détails requis** : décrivez les calculs utilisés pour déterminer le taux d'erreur aux variables non subjectives et donnez toute définition connexe. Lorsque vous évaluez ce critère, vous devez tenir compte de l'élément de données de base qui affiche le taux d'erreur le plus élevé. Le plus souvent, la quantité d'erreurs dans les éléments de données d'une base de données est évaluée grâce à des études de seconde saisie des données ou à d'autres études spéciales (et habituellement rétrospectives), mais il est aussi possible de le faire lors du processus d'élaboration de la base de données. Si le niveau d'erreur n'est pas estimé au moyen d'une étude de la qualité des données, cela ne veut pas nécessairement dire que le critère doit automatiquement recevoir la cote *inconnu*. Consultez l'annexe C pour de plus amples renseignements.

**Évaluation** :

- aucun ou minimal (taux d'erreur de 0 % à moins de 5 %)
- modéré (de 5 à 10 %)
- significatif (plus de 10 %)
- inconnu (le résultat n'a pu être déterminé)

**Critère 16** *Le niveau du biais n'est pas significatif.*

Évaluez ce critère ici.

**Détails requis** : expliquez s'il y a présence réelle ou présumée d'un biais important dans les données. Si un biais (ou un biais corrélé) réel ou présumé est suffisamment important pour influencer considérablement sur les estimations, vous devez attribuer la cote *non respecté* au niveau du biais. Dans le cas contraire, s'il n'y a aucun biais réel ou présumé, vous devez attribuer la cote *respecté*. Autrement, le critère devrait être coté comme *inconnu*.

**Évaluation** :

- respecté
- non respecté
- inconnu

**Critère 17** *Le degré de problèmes de cohérence se classe parmi l'une des catégories prédéterminées.*

Évaluez ce critère ici.

**Détails requis** : si la base de données contient des éléments de données qui dépendent de l'opinion ou de l'interprétation des codificateurs, donnez des renseignements sur la cohérence des mesures de chaque codificateur et sur la cohérence des mesures entre les codificateurs. Lorsque vous évaluez ce critère, vous devez tenir compte des éléments de données qui affichent le niveau de cohérence le plus faible.

**Évaluation** :  aucun ou minimal (taux d'erreur\* de 0 % à moins de 5 %)  
 modéré (de 5 à 10 %)  
 significatif (plus de 10 %)  
 inconnu (le résultat n'a pu être déterminé)

\* Vous pouvez également utiliser la statistique kappa; consultez l'annexe C pour plus de renseignements.

## 1.6 Vérification et imputation

**Critère 18** *Des contrôles de la validité sont réalisés pour chacun des éléments de données et toutes les données invalides sont signalées.*

Évaluez ce critère ici.

**Détails requis** : indiquez si des contrôles de la validité sont réalisés pour chacun des éléments de données, de même que le type de contrôle effectué, comme comparer la réponse à une liste de réponses acceptables ou vérifier que son format est approprié. Indiquez comment les données invalides peuvent être relevées et quels sont les processus de suivi (les données invalides sont exclues, renvoyées au fournisseur pour correction, signalées pour l'imputation ou simplement considérées comme invalides et traitées séparément).

**Évaluation** :  respecté  non respecté

**Critère 19** *Les règles de vérification et l'imputation sont logiques et appliquées de façon uniforme.*

Évaluez ce critère ici.

**Détails requis** : expliquez comment les règles de vérification et d'imputation sont appliquées aux données, y compris l'étape à laquelle elles sont appliquées (saisie des données, traitement des données).

**Évaluation** :  respecté  non respecté

**Critère 20** *Les rapports de vérification destinés aux utilisateurs sont faciles à utiliser et à comprendre.*

Évaluez ce critère ici.

**Détails requis** : décrivez l'approche utilisée par la banque de données pour produire des rapports de vérification destinés aux utilisateurs, la clarté des rapports et les actions requises de la part des destinataires des rapports de vérification. De plus, il pourrait être utile de préciser tout commentaire des utilisateurs sur les rapports de vérification.

**Évaluation** :  respecté  non respecté  non applicable

**Critère 21** *Le processus d'imputation est automatisé et conforme aux règles de vérification.*

Évaluez ce critère ici.

**Détails requis** : expliquez le processus d'imputation suivi par la banque de données, le cas échéant. Indiquez si l'imputation est automatisée et si elle obéit à des règles de vérification.

**Évaluation** :  respecté  non respecté  non applicable

## 1.7 Traitement et estimation

**Critère 22** *La documentation sur toutes les activités de traitement des données est tenue à jour.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur toutes les activités de traitement effectuées par le personnel de la banque de données, sur leur documentation et sur l'endroit où cette documentation peut être consultée.

**Évaluation** :  respecté  non respecté

**Critère 23** *Les spécifications techniques pour la banque de données sont tenues à jour.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur les systèmes, les programmes et les applications de la banque de données, sur leur documentation et sur l'endroit où cette documentation peut être consultée.

**Évaluation** :  respecté  non respecté

**Critère 24** *Les modifications à la structure sous-jacente ou aux programmes de traitement ou d'estimation d'une banque de données ont été testées.*

Évaluez ce critère ici.

**Détails requis** : indiquez tous les changements apportés aux bases de données sous-jacentes ou aux programmes d'estimation et de traitement de la banque de données au cours de la période de référence. Donnez également des détails sur les types de tests effectués (tests d'acceptation des unités, des systèmes ou des utilisateurs). Si aucune révision n'a eu lieu au cours de la dernière année, cette cote n'est pas applicable.

**Évaluation** :  respecté  non respecté  non applicable

**Critère 25** *Les données brutes, conformément à la politique de l'ICIS en matière de maintien des données, sont sauvegardées dans un endroit sécuritaire.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur l'endroit où sont stockées les données brutes de la banque de données et sur les procédures de stockage.

**Évaluation** :  respecté  non respecté

**Critère 26a** *Les statistiques agrégées d'une banque de données ont été comparées, dans la mesure du possible, à des statistiques similaires provenant d'une autre banque de données de l'ICIS ou d'une source externe.*

Évaluez ce critère ici.

**Détails requis** : donnez des détails, si possible, sur les comparaisons des données agrégées d'une banque de données et décrivez les résultats. Donnez des explications sur les comparaisons pour lesquelles il n'y a pas eu de correspondance parfaite. Portez une attention particulière aux divergences résultant de différences sur le plan des périodes de références, des populations de référence, des sources de données et du moment de la collecte des données.

**Évaluation** :  respecté  non respecté  non applicable

**NOUVEAU**

**Critère 26b** *La variance de l'estimation, comparée à l'estimation elle-même, se situe à un niveau acceptable.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur la valeur calculée pour le coefficient de variation, de même que sur le niveau acceptable pour les principaux clients de la banque de données. Ce critère s'applique uniquement aux estimations basées sur un échantillon. Les bases de données qui n'utilisent pas d'échantillons devraient coter ce critère comme *non applicable*. Il est important que les bases de données utilisant des données de recensement n'emploient pas le terme « variation » pour décrire la variabilité observée dans leurs données; elles doivent plutôt se servir d'autres mesures de la variabilité, comme l'écart, l'intervalle interquartile, l'écart moyen ou l'écart absolu moyen.

**Évaluation** :  respecté (coefficients de variation des estimations qui atteignent un niveau acceptable, moins de 16,6 %)  
 non respecté (le coefficient de variation est de 16,6 % ou plus ou n'est pas à un niveau acceptable pour les principaux clients de la banque de données)  
 non applicable

## 2 Dimension de l'actualité

### 2.1 Actualité des données au moment de leur diffusion

**Critère 27** *L'écart entre la date réelle de diffusion et la fin de la période de référence est raisonnablement bref.*

Évaluez ce critère ici.

**Détails requis** : indiquez l'écart entre la date réelle de diffusion et la fin de la période de référence. Donnez des renseignements sur les délais, s'il y a lieu, et précisez si les délais étaient causés par des événements ponctuels ou s'ils devraient se poursuivre. Dans le cas des bases de données qui n'ont pas de diffusion annuelle de données, toute diffusion de données aux principaux intervenants devrait servir de point de comparaison. Ce critère ne s'applique pas aux bases de données de nature longitudinale ni à celles qui n'ont pas de fin de période de référence.

**Évaluation** :  respecté  non respecté  non applicable

**Critère 28** *La date officielle de diffusion a été annoncée avant la diffusion.*

Évaluez ce critère ici.

**Détails requis** : indiquez si la date officielle de la diffusion annuelle des données ou de la diffusion de données aux principaux intervenants a été planifiée et annoncée au moins six mois à l'avance.

**Évaluation** :  respecté  non respecté  inconnu  non applicable

**Critère 29** *La diffusion a eu lieu à la date officielle prévue.*

Évaluez ce critère ici.

**Détails requis** : indiquez si les données ont été diffusées à la date officielle de diffusion ou avant. En ce qui concerne les diffusions de données aux principaux intervenants, l'évaluation de ce critère repose sur la comparaison entre la date de diffusion prévue (indiquée dans le calendrier des produits livrables) et la date de diffusion réelle.

Évaluation :  respecté  non respecté  non applicable

**Critère 30** *Les activités de traitement des données sont revues régulièrement aux fins d'amélioration de l'actualité des données.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur les programmes ou les systèmes utilisés pour préparer et analyser les données, de même que sur l'échéancier de révision périodique de ces activités et leur capacité à produire des données en temps opportun.

Évaluation :  respecté  non respecté

## 2.2 Actualité de la documentation

**Critère 31** *La documentation recommandée sur la qualité des données était disponible au moment de la diffusion des données ou des rapports.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur la disponibilité de la documentation sur la qualité des données à l'intention des utilisateurs une fois qu'ils ont accès, tant à l'interne qu'à l'externe, aux données ou aux rapports. Donnez des détails sur l'endroit où ces renseignements peuvent être consultés et sur la façon de les consulter. La section 4 présente les composantes qui devraient être incluses dans la documentation sur la qualité des données.

Évaluation :  respecté  non respecté

**Critère 32** *Les principaux rapports ont été diffusés à la date prévue.*

Évaluez ce critère ici.

**Détails requis** : indiquez si les principaux rapports de la banque de données ont été diffusés à la date prévue. S'il y a lieu, précisez les circonstances expliquant les retards, y compris les mesures prises pour aviser les utilisateurs.

Évaluation :  respecté  non respecté  non applicable

## 3 Dimension de la comparabilité

### 3.1 Normes du Dictionnaire de données

**Critère 33** *Tous les éléments de données font l'objet d'une évaluation visant à déterminer s'ils sont inclus dans le Dictionnaire de données de l'ICIS.*

Évaluez ce critère ici.

**Détails requis** : indiquez les résultats de la révision des éléments de données des banques de données existantes et de celles en cours d'élaboration par rapport au Dictionnaire de données de l'ICIS. Indiquez quels éléments ne correspondent pas aux normes du Dictionnaire de données ainsi que ceux auxquels ne correspond aucune norme dans le Dictionnaire de données de l'ICIS.

**Évaluation** :  respecté  non respecté  inconnu

**Critère 34** *Les éléments de données d'une banque de données qui sont inclus dans le Dictionnaire de données de l'ICIS doivent être conformes aux normes de ce dictionnaire.*

Évaluez ce critère ici.

**Détails requis** : pour les éléments de données figurant actuellement dans le Dictionnaire de données de l'ICIS, décrivez le niveau de conformité aux normes du dictionnaire ainsi que toute dérogation justifiable aux normes de l'ICIS. Dictionnaire de données ne comprend aucun élément de données de la banque de données, cette cote n'est pas applicable.

**Évaluation** :  respecté  non respecté  inconnu  non applicable

### 3.2 Normalisation

**Critère 35** *Les données sont recueillies au niveau de détail le plus élevé possible.*

Évaluez ce critère ici.

**Détails requis** : indiquez si tous les éléments de données de base sont recueillis avec le niveau de détail nécessaire aux fins de publication, de couplage ou de comparaison. Expliquez les cas où les données ne sont pas recueillies avec le niveau de détail le plus élevé possible.

**Évaluation** :  respecté  non respecté

**Critère 36** *Pour tout élément de données dérivé, l'élément de données original demeure accessible.*

Évaluez ce critère ici.

**Détails requis** : indiquez si les éléments de données originaux de la base de données sont présents et accessibles.

**Évaluation** :  respecté  non respecté  non applicable

### 3.3 Couplage

**Critère 37** *Les données géographiques sont recueillies au moyen de la Classification géographique type (CGT).*

Évaluez ce critère ici.

**Détails requis** : décrivez dans quelle mesure les entités sur lesquelles on recueille des données (établissements, personnes, provinces, etc.) sont identifiables soit par le code postal (de six caractères), soit par la classification géographique type pertinente. Si le plus bas niveau géographique utilisé est la province, vous devez utiliser les codes provinciaux standard de Postes Canada. Comme l'information géographique peut s'appliquer à plus d'une entité, les bases de données cliniques, par exemple, devraient recueillir de l'information géographique non seulement sur le patient, mais aussi sur l'établissement.

**Évaluation** :  respecté  non respecté

**Critère 38** *Les données sont recueillies selon le même calendrier, plus particulièrement d'une province ou d'un territoire à l'autre et au sein d'une province ou d'un territoire.*

Évaluez ce critère ici.

**Détails requis** : indiquez si l'information d'une province ou d'un territoire à l'autre et au sein d'une province ou d'un territoire est suffisante pour permettre l'analyse des données selon le même calendrier. Décrivez les difficultés ou les obstacles concernant ce type d'analyses.

**Évaluation** :  respecté  non respecté

**Critère 39** *L'utilisation d'identificateurs permet de distinguer les établissements ou les organismes seulement aux fins de couplage rétrospectif.*

Évaluez ce critère ici.

**Détails requis** : expliquez dans quelle mesure un code unique acceptable aux fins de couplage rétrospectif (identificateur assigné par la province ou le territoire, ou l'équivalent) existe pour chaque établissement ou organisme et est disponible dans la base de données.

**Évaluation** :  respecté  non respecté  non applicable

**Critère 40** *L'utilisation d'identificateurs permet de distinguer les personnes ou les appareils seulement aux fins de couplage rétrospectif.*

Évaluez ce critère ici.

**Détails requis** : indiquez si la base de données renferme un identificateur unique pour chaque personne ou appareil qui peut servir au couplage d'enregistrements correspondants sur différentes périodes de temps.

**Évaluation** :  respecté  non respecté  non applicable

### 3.4 Équivalence

**Critère 41** *La méthodologie et les limites des concordances ou des conversions sont documentées.*

Évaluez ce critère ici.

**Détails requis** : décrivez les concordances et les conversions utilisées par la banque de données, et indiquez à quelle fréquence leur méthodologie et leurs limites sont documentées pour les utilisateurs. Donnez le nom de ce document et l'endroit où il est possible d'y accéder, à l'interne et à l'externe.

**Évaluation** :  respecté  non respecté  non applicable

**Critère 42** *L'ampleur des problèmes liés aux concordances et aux conversions se classe parmi l'une des catégories prédéterminées.*

Évaluez ce critère ici.

**Détails requis** : évaluez l'efficacité des concordances et des conversions utilisées dans la base de données. Présentez les procédures d'essais des nouvelles concordances et conversions ainsi que les procédures permettant de relever et corriger les erreurs de classification. Si la base de données utilise plus d'une concordance ou conversion, faites l'évaluation globale à partir de celle dont la cote est la plus faible.

**Évaluation** :  minimal (peu ou pas de problèmes)  
 modéré (problèmes identifiables d'envergure limitée)  
 significatif (les données sont en grande partie mal converties, ce qui influe sur les résultats)  
 inconnu (l'équivalence n'a pas été examinée)  
 non applicable

### 3.5 Comparabilité rétrospective

**Critère 43** *La documentation des modifications antérieures apportées à la banque de données existe et est facilement accessible.*

Évaluez ce critère ici.

**Détails requis** : indiquez s'il existe une documentation des modifications antérieures et si elle est groupée en un document. Précisez la fréquence de sa mise à jour. Donnez des détails sur le contenu du document, qui devrait inclure les changements apportés aux concepts, aux méthodologies, aux bases de sondage et aux éléments de données.

**Évaluation** :  respecté  non respecté

**Critère 44** *La réalisation d'une analyse des tendances permet d'examiner les changements dans les éléments de données de base au fil du temps.*

Évaluez ce critère ici.

**Détails requis** : indiquez si une analyse des tendances a été réalisée pour les éléments de données de base depuis la dernière évaluation de la qualité des données.

**Évaluation** :  respecté  non respecté  non applicable

**Critère 45** *L'ampleur des problèmes découlant de la comparaison des données au fil du temps se classe parmi l'une des catégories prédéterminées.*

Évaluez ce critère ici.

**Détails requis** : donnez des détails sur les limites et les difficultés inhérentes à la production d'estimations tendanciennes valides.

**Évaluation** :  minimal (peu ou pas de problèmes dans la production de tendances comparables)  
 modéré (problèmes décelés dans quelques données tendanciennes)  
 significatif (des données tendanciennes exactes ne peuvent pas être produites pour un élément de données de base)  
 inconnu (il n'a pu être déterminé si des tendances exactes peuvent être produites)  
 non applicable

## 4 Dimension de la facilité d'utilisation

### 4.1 Accessibilité

**Critère 46** *Un ensemble de données final est rendu disponible à la date prévue de diffusion.*

Évaluez ce critère ici.

**Détails requis** : indiquez si les données utilisées pour l'analyse et la création de rapports ou de diffusions sont sauvegardées dans un endroit sécuritaire pour consultation future. Donnez des détails sur le format des données et sur la façon par laquelle il est possible d'y accéder.

**Évaluation** :  respecté  non respecté

**Critère 47** *Des tableaux et des analyses au format et au contenu standard sont produits lors de chaque diffusion prévue ou sur demande.*

Évaluez ce critère ici.

**Détails requis** : donnez des détails sur les analyses et les tableaux standard généralement utilisés, y compris sur la façon dont les utilisateurs peuvent y accéder pour chaque diffusion prévue.

**Évaluation** :  respecté  non respecté

**Critère 48** *Les produits sont définis, catalogués ou annoncés.*

Évaluez ce critère ici.

**Détails requis** : décrivez les systèmes de diffusion de l'ICIS utilisés pour définir, cataloguer ou annoncer les produits de la banque de données pour chaque diffusion prévue.

**Évaluation** :  respecté  non respecté

### 4.2 Documentation

**Critère 49** *Il existe de la documentation sur la qualité des données à l'intention des utilisateurs.*

Évaluez ce critère ici.

**Détails requis** : indiquez si un document indépendant sur la qualité des données est mis à la disposition des utilisateurs (internes et externes) au moins une fois par année. Précisez le format du document, l'endroit où il se trouve et son calendrier de mises à jour.

**Évaluation** :  respecté  non respecté

**Critère 50** *Il existe de la documentation sur les métadonnées.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur la documentation sur les métadonnées de la banque de données aux fins d'utilisation interne et sur les pratiques de révision et de mise à jour de la documentation. Veuillez préciser où se trouve la documentation interne sur les métadonnées.

**Évaluation** :  respecté  non respecté

**Critère 51** *Un avertissement accompagne toute diffusion préliminaire.*

Évaluez ce critère ici.

**Détails requis** : indiquez si la banque de données a fourni des diffusions préliminaires de données officielles ou non officielles pour la période en question. Précisez si un avertissement a été émis et fournissez des détails sur le contenu de cet avertissement.

**Évaluation** :  respecté  non respecté  non applicable

### 4.3 Interprétabilité

**Critère 52** *Un mécanisme permet aux principaux utilisateurs de communiquer leurs commentaires à la section d'une banque de données et de recevoir une réponse de cette dernière.*

Évaluez ce critère ici.

**Détails requis** : présentez les pratiques de la banque de données qui permettent aux utilisateurs (internes et externes) de faire part de leur rétroaction à la section de l'ICIS ainsi que les mécanismes qui facilitent les communications entre l'ICIS et les principaux utilisateurs.

**Évaluation** :  respecté  non respecté

**Critère 53** *Des lignes directrices de révision sont disponibles et appliquées à chaque diffusion.*

Évaluez ce critère ici.

**Détails requis** : indiquez si des lignes directrices propres à une banque de données sont disponibles et appliquées chaque fois que des données sont diffusées ou extraites. Prenez note que le calendrier de révision de ces lignes directrices doit être à jour et indiquez où les lignes directrices se trouvent.

**Évaluation** :  respecté  non respecté

## 5 Dimension de la pertinence

### 5.1 Adaptabilité

**Critère 54** *Des mécanismes tiennent les clients et les intervenants informés des développements dans le domaine.*

Évaluez ce critère ici.

**Détails requis** : décrivez les mécanismes de liaison mis en place par le personnel de la section de la banque de données pour informer les intervenants des développements dans le domaine.

**Évaluation** :  respecté  non respecté

**Critère 55** *La banque de données est conçue de manière à faciliter les modifications futures du système.*

Évaluez ce critère ici.

**Détails requis** : évaluez la facilité d'intégration des modifications futures en indiquant si la banque de données a démontré sa capacité à s'adapter à une importante question émergente, à une nouvelle norme technique ou à une limite marquée de la qualité des données au cours de l'année précédente.

**Évaluation** :  respecté  non respecté

### 5.2 Valeur

**Critère 56** *Le mandat de la banque de données comble une lacune en matière d'information sur la santé.*

Évaluez ce critère ici.

**Détails requis** : donnez des renseignements sur le mandat de la banque de données et la lacune qu'elle comble dans le secteur des soins de santé. Indiquez également si la banque de données a récemment été évaluée en fonction des autres banques de données à l'ICIS et dans tout le secteur externe.

**Évaluation** :  respecté  non respecté

**Critère 57** *Le niveau d'utilisation de la banque de données est surveillé.*

Évaluez ce critère ici.

**Détails requis** : énumérez les approches employées par le personnel de la banque de données pour surveiller l'utilisation de la banque (p. ex. surveiller le nombre et le type de demandes de données, l'utilisation des rapports électroniques, les utilisations de données qui retiennent l'attention du public, les visites des pages Web, les coupures de presse, les annonces, les citations, les ouvrages rédigés par le personnel, les ventes, les communications avec le personnel et l'apparition du personnel dans les médias, les conférences et les forums stratégiques). Donnez des détails sur une utilisation des données particulièrement intéressante ou sur le type de données le plus fréquemment demandé.

**Évaluation** :  respecté  non respecté

**Critère 58** *La satisfaction des utilisateurs est évaluée périodiquement.*

Évaluez ce critère ici.

**Détails requis** : décrivez les méthodes utilisées pour évaluer la satisfaction des clients et la fréquence d'utilisation. Indiquez le degré de satisfaction général rapporté par les utilisateurs, les intervenants et le personnel de l'ICIS.

**Évaluation** :  respecté  non respecté



## Annexe E – Sous-catégories pour la documentation sur les métadonnées

**NOUVEAU**

Comme le mentionne la section 4.3, il existe sept catégories dans la documentation sur les métadonnées de l'ICIS pour les banques de données. Chacune de ces catégories est divisée en sous-catégories.

### Catégorie : description de la base de données

Sous-catégorie	Définition ou exemple
Contexte	Information donnant aux utilisateurs un aperçu général de la base de données (p. ex. base de données nationale sur tous les hôpitaux de soins de courte durée).
But	Objectif de la base de données (p. ex. recueillir des données sur les congés des patients hospitalisés).
Portée	Sujets d'enquête et d'analyse qui intéressent les utilisateurs (p. ex. toutes les provinces à l'exception du Québec).
Intervenants	Entités directement intéressées par les résultats d'une base de données (p. ex. ministère de la Santé et des Soins de longue durée de l'Ontario).
Rôles et responsabilités des différents membres de l'équipe du projet, y compris les fournisseurs de données	Description (générique, si possible) du rôle de chaque membre de l'équipe de projet (p. ex. les fournisseurs de données transmettent à l'ICIS les données les plus récentes avant une certaine date et dans un format précis).
Définition des concepts	Description des termes propres à une base de données (p. ex. sortie d'un hôpital = congé ou décès d'un patient hospitalisé).
Liste des personnes-ressources internes	Membres du personnel de l'ICIS avec qui il est possible de communiquer pour obtenir plus de renseignements sur une base de données (coordonnateur, analyste principal, etc.).

## Catégorie : méthodologie

### Méthodologie — population cible et population de référence

Sous-catégorie	Définition ou exemple
Définition	Description de la population pour laquelle on désire obtenir de l'information et population disponible.
Exclusions	Description des enregistrements qui sont exclus de la population cible pour une raison ou une autre.

### Méthodologie — base de sondage

Sous-catégorie	Définition ou exemple
Définition	Liste des unités qui donnent accès à la population de référence (p. ex. établissements comptant des lits d'hospitalisation en soins de courte durée au Canada, sauf le Québec).
Liste des éléments de données disponibles	Quels sont les éléments de données disponibles dans les enregistrements de la base de sondage?
Liste des unités et des personnes-ressources	Quelles unités doivent se trouver dans la base de sondage et quelles sont les coordonnées des fournisseurs de données en cas de questions ou de préoccupations?
Modifications antérieures à la base de sondage	Description de l'évolution de la base de sondage depuis sa conception (p. ex. nombre d'enregistrements, changements aux fournisseurs de données, changements à la couverture).
Mise à jour de la base de sondage	
– ajout, retrait et fusion d'unités	Procédure en place pour traiter les unités de la base de sondage en cas d'ajouts, de suppressions, de fusions, etc.
– mise à jour des unités	Procédure en place pour traiter les éléments de données ou les enregistrements qui ont changé.

## Méthodologie – modèle du sondage (sondage seulement)

Sous-catégorie	Définition ou exemple
Description du plan du sondage ou explication (stratégie d'échantillonnage, méthode utilisée)	Description de la raison du choix d'un échantillon et de la façon dont il a été choisi (c'est-à-dire économies de temps et d'argent, échantillon aléatoire dans chaque province, éléments de données de la stratification).
Définition des unités d'échantillonnage	Description de ce qui est considéré comme une unité d'échantillonnage, par exemple établissement, dossier (unités distinctes qui ne se chevauchent pas, dans lesquelles la population est divisée en vue de la sélection d'un échantillon).
Choix d'un échantillon	
– calcul de la taille de l'échantillon	Description de la manière dont la taille de l'échantillon a été déterminée.
– répartition de l'échantillon	Description de la répartition de l'échantillon dans la strate (p. ex. 100 dossiers par province).
Probabilité de sélection	Proportion du nombre d'unités échantillonnées par rapport au nombre d'unités de population dans chaque strate.

## Catégorie : collecte et saisie de données

Sous-catégorie	Définition ou exemple
Liste d'éléments de données	
– valeurs acceptées	Vérifications de la validité qui existent pour chaque élément de données saisi.
– modifications antérieures	Documentation présentant les éléments de données dont les valeurs possibles ont changé au fil du temps.
– disposition de l'enregistrement	Disposition prévue des enregistrements contenus dans le fichier de données reçu du fournisseur de données.
– changements aux éléments de données	Documentation présentant les éléments de données qui ont changé au fil du temps (ajouts, suppressions, ajouts de détails, suppressions de détails).
– Inventaire du Dictionnaire de données	Contient les noms et les définitions des éléments de données, ainsi que les spécifications de chacun d'entre eux.

<b>Sous-catégorie</b>	<b>Définition ou exemple</b>
<b>Spécifications du fournisseur ou de l'application</b>	
– disposition de l'enregistrement	Disposition de l'enregistrement exigée pour le traitement du fichier de données du fournisseur.
– changements aux spécifications	Documentation des changements effectués au fil des ans concernant les exigences du système de traitement à l'égard du fichier de données.
– liste des fournisseurs approuvés (le cas échéant)	Liste des fournisseurs actuellement utilisés, des fournisseurs recommandés à la suite d'anciens projets et des fournisseurs considérés pour des projets futurs.
<b>Coordonnées des clients</b>	
– procédures relatives à la communication avec les clients	Procédures relatives à la communication avec les clients pour vérifier et mettre à jour les renseignements.
– coordonnées	Liste des renseignements requis pour communiquer avec les clients (nom, titre de poste, numéro de téléphone, courriel).
– lettre dans laquelle figurent les coordonnées du fournisseur de données	Lettre contenant les coordonnées de la personne-ressource du fournisseur de données.
– registre des communications	Document détaillant toutes les communications avec les clients (méthode de communication, avec qui a-t-on communiqué et à quel moment, suivi, etc.).
– suivi à la suite d'une non-réponse	Procédure à suivre (lettre, appel téléphonique, etc.) lorsque les données attendues (au niveau de l'enregistrement ou de l'élément de données) ne sont pas reçues.
<b>Spécifications pour la soumission des données</b>	
– échéance	Date prévue de la soumission de données du fournisseur.
– convention de désignation des fichiers	Convention de désignation des noms des fichiers de données provenant du fournisseur.
– disposition détaillée de l'enregistrement	Disposition prévue des fichiers provenant des fournisseurs de données.
– méthode de collecte des données	Façon dont seront soumises les données par le fournisseur de données (en ligne, par courriel, sur CD ou disquette, etc.).

<b>Sous-catégorie</b>	<b>Définition ou exemple</b>
Manuel de codification de la base de données	
– copie de l'abrégé, du formulaire, du questionnaire	Le manuel doit inclure une copie des questionnaires et des formulaires utilisés par le fournisseur de données pour recueillir des renseignements.
– traduction	Le manuel doit inclure tous les documents traduits.
– instruction aux fournisseurs de données	Le manuel doit inclure des instructions sur la façon dont les fournisseurs doivent soumettre les données.
– liste et description des valeurs acceptables pour les éléments de données obligatoires et facultatifs	Le manuel doit indiquer quels sont les éléments de données obligatoires et quelles valeurs sont acceptables.
– liste des normes de codification (le cas échéant)	Le manuel doit clairement indiquer quelles normes de codification doivent être suivies par le fournisseur de données (p. ex. CIM-9, CIM-10-CA/CCI).
Normalisation des éléments de données provenant de différents fournisseurs de données	
– procédures de normalisation	Procédures de normalisation des valeurs qui n'étaient pas normalisées lors de leur soumission.
Consolidation des données dans un fichier commun	
– processus de fermeture de la base de données	Procédure permettant de créer un seul fichier de données à partir des différents fichiers reçus des fournisseurs de données.

## Catégorie : traitement des données

Sous-catégorie	Définition ou exemple
Diagramme du cheminement des données	
– programmes	Documentation ou diagramme décrivant les programmes de traitement des données.
– fichiers d'entrée et de sortie	Documentation ou diagramme décrivant les fichiers d'entrée requis pour les programmes de traitement des données ainsi que les fichiers de sortie qui en résultent.
Vérification	
– liste des codes valides pour chaque élément de données	Liste des codes valides pour les éléments de données à chaque étape du traitement des données.
– contrôles de la validité	Documentation de tous les contrôles de la validité effectués à chaque étape du traitement des données.
– contrôles de la cohérence	Documentation de tous les contrôles de vérification de la cohérence et de la répartition effectués à chaque étape du traitement des données.
– contrôles de détection des valeurs aberrantes	Documentation de tous les contrôles de détection des valeurs aberrantes effectués à chaque étape du traitement des données.
– registre de tous les rejets à la vérification	Fichier contenant les enregistrements ou les éléments de données qui ont échoué aux différents contrôles (renseignements permettant d'identifier l'enregistrement, renseignements décrivant les erreurs).
– exigences minimales relatives à une réponse	Critères en place permettant de s'assurer que la qualité n'est pas compromise en raison de réponses invalides ou incomplètes (éléments exclus, non-réponse totale, réponse partielle, réponse complète).
– mesures prises pour les rejets à la vérification signalés	Procédure pour le traitement des rejets à la vérification et des données manquantes (communiquer avec les fournisseurs de données, imputation, repondération).
– processus de vérification	Documentation du processus de vérification complet.
Codification (questions ouvertes, questions fermées avec la catégorie « autre »)	Procédure pour l'attribution de codes aux réponses qui résultent de questions ouvertes ou fermées.

<b>Sous-catégorie</b>	<b>Définition ou exemple</b>
<b>Imputation</b>	
– description des méthodes d'imputation	Description (en cas d'imputation) des méthodes utilisées pour remplacer les valeurs invalides ou manquantes.
– champs	Éléments de données pouvant être imputés ou ayant été imputés.
– création d'éléments de données imputés	Description du calcul des valeurs au moyen de la méthode d'imputation pour remplacer les valeurs invalides ou manquantes.
– registre des imputations	Fichier contenant les enregistrements ou les éléments de données qui renferment les valeurs imputées, ainsi que la méthode utilisée (si plus d'une méthode a pu être utilisée).
– processus d'imputation	Documentation du processus d'imputation complet.
<b>Création d'éléments de données dérivés et groupés</b>	
– description	Documentation sur les éléments de données groupés ou dérivés à partir d'autres éléments ainsi que sur les raisons de ces groupements ou dérivations.
– noms des éléments de données	Documentation sur les noms des éléments de données obtenus par dérivation ou groupement.
– source	Documentation sur les éléments de données qui ont servi de base à la dérivation ou au groupement.
– calcul	Documentation détaillée sur la transformation de données originales en éléments de données dérivés ou groupés.
– registre des modifications antérieures	Fichier contenant l'ensemble des dérivations et des groupements d'éléments de données qui ont lieu au fil des ans.
– processus de création d'éléments de données dérivés	Documentation sur le processus utilisé pour la dérivation ou le groupement des éléments de données.
– processus pour le changement du processus de dérivation	Documentation sur le processus à suivre si des changements doivent être apportés à la dérivation ou au groupement.

## Traitement des données — pondération (sondage seulement)

Sous-catégorie	Définition ou exemple
Méthode de pondération :	
– poids d'échantillonnage	Liste des poids d'échantillonnage pour chaque enregistrement échantillonné (nombre d'unités dans la population que représente chaque unité échantillonnée).
– réglages de la pondération	Processus utilisé pour ajuster les pondérations et ainsi tenir compte des non-réponses, améliorer les estimations, etc.
– pondération finale	Liste des pondérations finales pour chaque enregistrement échantillonné (combinaison de la pondération de l'échantillonnage et du réglage de la pondération).
Processus de pondération	Documentation sur la façon dont toutes les pondérations sont calculées.

## Traitement des données — production du fichier analytique

Sous-catégorie	Définition ou exemple
Création des fichiers (ou du fichier des totalisations dans le cas d'un sondage)	Documentation des étapes à suivre pour produire le fichier qui sera utilisé dans le calcul des estimations (étapes à suivre, structure et couplage, format de fichiers).
Emplacement	Documentation précisant l'endroit où se trouve le fichier qui sera utilisé dans le calcul des estimations.
Disposition détaillée des enregistrements (éléments de données recueillis et dérivés)	Documentation précisant la disposition de l'enregistrement du fichier qui sera utilisé dans le calcul des estimations.
Exemple de lettres et de documents (CD, disquette, fichier non hiérarchique) envoyés aux fournisseurs de données pour révision et validation	Lettres ou documents envoyés aux fournisseurs de données afin qu'ils revoient les estimations obtenues à partir d'une analyse.
Accès aux données (qui, comment, quel type d'accès)	Protocole d'accès aux estimations qui résultent des données.
Description ou explication des méthodes d'estimation et des méthodes d'estimation de la variance utilisées (sondage seulement)	Documentation de la méthodologie utilisée pour calculer les estimations et les variances.
Processus de création des estimations	Documentation du calcul des estimations ou des résultats.

## Catégorie : Analyse et diffusion des données

Sous-catégorie	Définition ou exemple
Production de tableaux	
– liste des tableaux couramment produits	Documentation sur les différents tableaux de données de sortie produits (pour publication, etc.).
– modèles de tableaux	Document contenant les modèles des différents tableaux produits.
– tableaux rétrospectifs	Document contenant les tableaux qui ont été produits au fil des ans (aux fins de comparaison, pour des demandes spéciales, etc.).
– programmes visant la production de tableaux	Document indiquant l'emplacement des programmes et leur utilité dans la création de tableaux.
– processus de création de tableaux	Document à jour décrivant le processus complet de création de tableaux.
– processus de validation des tableaux	Document énumérant les différents contrôles de la validation utilisés pour s'assurer que les tableaux contiennent des réponses acceptables.
Suppression de l'information de nature confidentielle	
– méthodes de suppressions utilisées	Description de la méthode de suppression utilisée (suppression des cellules, codification en intervalles, regroupement des valeurs extrêmes supérieures et inférieures, arrondissement).
– critère de suppression	Description des critères entraînant la suppression d'une cellule s'ils ne sont pas respectés (p. ex. trop peu de réponses).
– chiffrement de données	Description de la méthode de chiffrement utilisée (p. ex. pour les numéros d'assurance-maladie).
– processus de suppression	Documentation du processus complet de suppression de cellules.

<b>Sous-catégorie</b>	<b>Définition ou exemple</b>
<b>Création de rapports</b>	
– modèle des rapports couramment produits	Document contenant les modèles des différents rapports produits.
– rapports précédemment produits	Document contenant les rapports qui ont été produits au fil des ans (aux fins de comparaison, etc.).
– processus de création des rapports	Documentation du processus complet de création de rapports.
– processus de validation des rapports	Document énumérant les différents contrôles de la validation utilisés pour s'assurer que le contenu des rapports est adéquat.
– approbation par les fournisseurs de données	Documentation qui sert à s'assurer que les fournisseurs de données ont approuvé tous les rapports et qui explique le processus d'approbation.
– processus de diffusion des données	Documentation du processus de diffusion des données.
<b>Publications</b>	
– liste et description des produits et services disponibles	Liste détaillée des publications et des services offerts aux utilisateurs.
– notes méthodologiques	Inclusion d'une section de notes méthodologiques (y compris une section sur la qualité des données) dans la publication.
– avis relatif aux données sur l'échantillon (y compris les coordonnées des personnes-ressources)	La publication indique à quel moment les estimations préliminaires ont été diffusées.
– foire aux questions	La publication contient une foire aux questions qui fournit une aide rapide aux utilisateurs.
– processus de publication	Documentation du processus complet de publication (avec qui communiquer, etc.).

Sous-catégorie	Définition ou exemple
Demandes spéciales	
– formulaire de demande de données	Un formulaire de demande de données existe pour les futures demandes des utilisateurs.
– liste des demandes	Une liste des demandes traitées existe.
– demandes sauvegardées	Une liste des demandes non traitées existe et est utilisée pour répondre aux demandes.
– programmes utilisés pour générer les résultats	Documentation des programmes utilisés pour répondre aux demandes de données.
– processus relatif aux demandes de données	Documentation du processus complet de réponse aux demandes de données (qui est responsable, date butoir, etc.).
– processus de validation des demandes	Documentation de la validation des données résultant de la demande avant de les mettre à la disposition des clients (vérification indépendante).

## Catégorie : stockage des données

Sous-catégorie	Définition ou exemple
Exigences relatives au stockage des données	Documentation sur les exigences relatives au stockage des données.
Fichiers rétrospectifs (emplacement et disposition de l'enregistrement)	Documentation sur l'emplacement et le contenu des fichiers rétrospectifs.
Inventaire du Dictionnaire de données	Contient les noms et les définitions des éléments de données, ainsi que les spécifications de chacun d'entre eux.

## Catégorie : documentation

Sous-catégorie	Définition ou exemple
Notes chronologiques	
– limites des données rétrospectives	Documentation sur les limites des données rétrospectives.
– comparabilité avec d'autres sources	Documentation concernant la comparabilité avec d'autres sources.
– changements à la méthodologie, à la collecte et au traitement	Documentation décrivant tout changement apporté à la méthodologie, à la collecte et au traitement des données au fil des ans.

<b>Sous-catégorie</b>	<b>Définition ou exemple</b>
Documentation propre aux différents utilisateurs avec des niveaux de détail différents	Différents types de documents existent pour différents types d'utilisateurs (analystes, chercheurs). Inclut les rapports des études de seconde saisie et d'autres études spéciales de la Division de la qualité des données ainsi que les rapports d'évaluation de la qualité de données qui en résultent.
Processus de création de documents pour différents utilisateurs	Il existe un processus pour la production de documents destinés à des utilisateurs particuliers (qui en est responsable, à quel moment les documents doivent-ils être rédigés, validation du rapport, personne-ressource de la Division de la qualité des données, etc.).
Rapport d'évaluation de la qualité des données	Un rapport d'évaluation de la qualité des données a été produit pour la dernière période de déclaration.
Processus de création d'un rapport d'évaluation de la qualité des données	Un processus de création d'un rapport d'évaluation de la qualité des données a été établi (qui en est responsable, à quel moment le rapport doit-il être rédigé, validation du rapport, personne-ressource de la Division de la qualité des données, etc.).
Documentation sur la qualité des données à l'intention des utilisateurs externes	Une documentation sur la qualité des données à l'intention des utilisateurs externes est produite. Inclut les rapports des études de seconde saisie et d'autres études spéciales de la Division de la qualité des données ainsi que les rapports d'évaluation de la qualité de données qui en résultent.
Processus de création d'une documentation sur la qualité des données à l'intention des utilisateurs externes	Un processus de création d'une documentation sur la qualité des données à l'intention des utilisateurs externes a été établi (qui en est responsable, quelle est la date butoir, validation, etc.).
Séances d'information et ateliers	Des séances d'information ou des ateliers sont organisés pour les clients externes ou le personnel de l'ICIS. Ces séances et ateliers les renseignent sur les banques de données et répondent à leurs questions.
Diagramme de Gantt — étapes, dates de début et de fin, personne responsable	Outil utilisé pour s'assurer que les activités requises se déroulent selon l'horaire établi et sont terminées à temps, ainsi que pour indiquer qui en est responsable.
Processus de création et de mise à jour du diagramme de Gantt	Il existe un processus permettant de créer et de mettre à jour le diagramme de Gantt afin qu'il demeure pertinent (qui en est responsable, à quelle fréquence, communication avec l'équipe de projet).

## Annexe F — Glossaire

**abrégé** : Récapitulation de l'information provenant d'un dossier clinique. Le processus qui consiste à récapituler les données est appelé *saisie de l'information*.

**accessibilité** : Une caractéristique. Facilité avec laquelle on peut obtenir des données d'une banque de données de l'ICIS.

**actualité** : Une dimension. Mesure le degré d'actualité et de mise à jour des données au moment de leur diffusion. L'actualité des données se mesure par l'écart entre la fin de la période de référence à laquelle les données se rapportent et la date à laquelle les données deviennent accessibles aux utilisateurs.

**actualité de la documentation** : Une caractéristique. Les critères qui se rapportent à cette caractéristique évaluent si la documentation sur la qualité des données recommandée et tout autre rapport important de la banque de données ont été rendus disponibles selon les besoins ou les plans.

**actualité des données** : Une caractéristique. L'actualité désigne principalement le caractère courant ou à jour des données au moment de leur diffusion. Elle se mesure par l'intervalle entre la date de diffusion et la dernière date à laquelle les données se rapportent.

**adaptabilité** : Une caractéristique. Mesure dans laquelle la situation et la souplesse d'une banque de données répondent aux besoins en information actuels et futurs de ses principaux utilisateurs.

**ajustement de courbe** : Processus d'ajustement d'une fonction curvilinéaire (ou courbe) à des points de données. Une fonction curvilinéaire est celle dont la valeur, lorsqu'elle est inscrite, suivra une ligne continue (mais pas nécessairement droite) — comme une courbe polynomiale, logistique, exponentielle ou sinusoïdale.

**attributs des données** : Caractéristiques d'un élément de données, par exemple le nom de l'élément de données, le domaine de valeurs, le type et l'étendue des données.

**aucun** : Une cote. Absence de l'attribut mesuré.

**base de données administratives** : Base de données dont l'information est principalement recueillie pour la tenue de registres, les finances ou des fins autres que la recherche.

**base de sondage** : Liste de toutes les unités (p. ex. provinces, établissements ou médecins) qui permet de s'assurer que l'ensemble des unités de la population de référence sont recueillies. Par exemple, la base de sondage d'une base de données administratives peut servir à déterminer la proportion des données qui ont véritablement été reçues.

**biais** : Évaluation des différences systématiques entre les valeurs déclarées d'une banque de données et les valeurs qui *auraient dû* être déclarées. Indique si les erreurs présentes se sont produites de façon aléatoire ou non. Voir aussi *biais corrélé*.

**biais corrélé** : Erreur systématique dans un élément de données associé à un autre élément de données de la base de données.

**caractéristique** : Aspect de la qualité des données composé d'un critère ou plus.

**CCI** : Classification canadienne des interventions en santé.

**cheminement des données** : Chemin que parcourent les données pendant leur traitement. Souvent décrit au moyen d'un diagramme illustrant les diverses sources de données, les étapes de traitement et les résultats.

**CIM-9** : Classification statistique internationale des maladies et des problèmes de santé connexes, 9<sup>e</sup> révision. Ensemble de codes acceptés à l'échelle internationale pour la classification des états et des diagnostics médicaux.

**CIM-10-CA** : Version canadienne enrichie de la CIM-10 de l'Organisation mondiale de la Santé. Elle se nomme Classification statistique internationale des maladies et des problèmes de santé connexe, 10<sup>e</sup> révision, Canada.

**Classification géographique type (CGT)** : Système de classification hiérarchique défini par Statistique Canada, qui regroupe différents secteurs géographiques. Les plus petits secteurs d'agrégat sont les secteurs de dénombrement, qui sont à leur tour nichés progressivement dans des groupements plus vastes, comme le secteur de recensement et la division de recensement, pour enfin aboutir à la province et au pays. Cette classification se fait souvent au moyen d'un fichier de conversion des codes postaux.

**coefficient de variation** : Calcul statistique mesurant la variation relative d'une distribution qui divise l'écart-type par l'estimation. On l'exprime souvent par un pourcentage.

**cohérence** : Mesure de la variation dans les réponses après des mesures répétées. Désignée aussi comme la fiabilité.

**collecte** : Voir *collecte des données*.

**collecte des données** : Rassemblement des données présentées par différents fournisseurs de données dans une banque de données commune.

**comparabilité** : Une dimension. Mesure avec laquelle des bases de données sont uniformes dans le temps et utilisent des conventions standard, comme des périodes de déclaration ou des éléments de données standard, ce qui les rend semblables à d'autres bases de données.

**comparabilité rétrospective** : Une caractéristique. Mesure de l'uniformité des concepts et des méthodes des données au fil du temps qui permet de faire des comparaisons valides pour différentes périodes.

**conventions standard** : Ensemble de processus ou d'attributs habituels, par exemple éléments de données ou périodes de déclaration standard.

**couplage** : Une caractéristique. Mesure la facilité avec laquelle une banque de données peut être appariée à une autre.

**couplage d'enregistrements** : Processus de jumelage des enregistrements provenant de deux bases de données ou plus, à l'aide d'un ou plusieurs éléments de données de liaison communs. Les éléments de données de liaison devraient posséder les mêmes attributs (comme le nom, l'étendue, le type ou le format des éléments de liaison).

**couverture** : Une caractéristique. Mesure dans laquelle la base de sondage d'une base de données décrit la population de référence.

**critère** : Énoncé spécifique qui se réfère à un élément détaillé de la qualité des données. Chaque critère reçoit une cote, soit *respecté*, *non respecté*, *inconnu* ou *non applicable*. Dans des cas précis, les critères reçoivent la cote *minimal* ou *aucun*, *modéré*, *significatif* ou *inconnu* en fonction d'autres catégories prédéterminées.

**cycle de production** : Processus utilisés dans la production d'un ensemble régulier de rapports ou de produits livrables à partir de la collecte de données jusqu'au traitement et pendant la création des produits finaux.

**cycle de travail de la qualité des données** : Approche en trois volets appliquée à la qualité des données, comprenant un ensemble d'activités de planification, de mise en œuvre et d'évaluation.

**cote de diffusion** : Date officielle à laquelle un ensemble annuel de données provenant d'une banque de données devient accessible aux utilisateurs.

**cépersonnaliser** : Supprimer des données toute information permettant aux utilisateurs d'identifier une personne, un établissement ou un équipement médical.

**Dictionnaire de données de l'ICIS** : Le Dictionnaire de données renferme les éléments et les définitions approuvés par l'équipe interne du dictionnaire. Le nom des éléments est conforme au modèle conceptuel des données sur la santé et, lorsque c'est possible, aux normes internationales comme HL7 et ISO.

**ciffusion** : Tout rapport, toute publication de données ou tout produit provenant d'une banque de données.

**diffusion annuelle** : Ensemble de données standard fourni annuellement aux utilisateurs des données. Cet ensemble peut comprendre, pour une période de référence donnée, une base de données complète ou un sous-ensemble de données.

**diffusion préliminaire non officielle** : Toute diffusion préliminaire d'un sous-ensemble de données, conçue pour authentifier ou valider des données. Par exemple, avant leur publication, on pourrait envoyer des chiffres sur les indicateurs de santé aux régions sanitaires aux fins de vérification.

**diffusion préliminaire officielle** : Diffusion d'un sous-ensemble de données peut-être incomplet dans le but d'améliorer l'actualité. Par exemple, des données recueillies sur une base annuelle pourraient être diffusées six mois avant la fin de l'année afin que les planificateurs du système de santé aient un premier aperçu des données complètes qui suivront.

**dimension** : Composantes distinctes de la définition plus vaste de la qualité des données.

**dispensateur de soins de santé** : Personne qui a offert, qui offre ou qui pourrait offrir des services ou des biens liés aux soins de santé. Au moment de la rédaction de la présente version du cadre de la qualité des données de l'ICIS, il s'agissait du seul concept final du Dictionnaire de données de l'ICIS.

**Document principal des méthodes** : Source exhaustive et à jour renfermant toute l'information détaillée au sujet d'une banque de données et les processus de collecte et de traitement utilisés.

**documentation** : Une caractéristique. Information sur la qualité des données, les méthodes et les avertissements d'une banque de données.

**documentation sur la qualité des données à l'intention des utilisateurs** : Documents fournissant de l'information suffisante pour que les utilisateurs déterminent si la qualité des données est appropriée pour l'utilisation qui en est prévue. Sert principalement à mettre en évidence les méthodes utilisées pour la collecte et la manipulation des données et à indiquer les principales limites des données.

**domaine des valeurs** : Fourchette des valeurs permises dans un élément de données.

**données brutes** : Données non modifiées provenant d'un fournisseur de données.

**données longitudinales** : Données qui, s'étendant sur un intervalle de temps, peuvent être utilisées pour évaluer des variations au fil du temps (p. ex. information sur les honoraires d'un médecin pendant 10 ans).

**double saisie** : Acte de recueillir les mêmes données deux fois et de les comparer pour réduire les erreurs de saisie et de soumission.

**élément de données de base** : Tout élément de données d'une banque de données qui est couramment utilisé dans une analyse.

**élément de données de liaison** : Élément de données commun à deux ensembles de données ou plus et qui peut être utilisé pour les combiner.

**élément de données dérivé** : Élément de données obtenu à partir d'autres éléments de données.

**enregistrements hors du champ d'observation** : Enregistrements qui ne devraient pas être inclus dans la banque de données. Comprend les enregistrements reçus d'unités à l'extérieur de la population de référence. Voir aussi *surdénombrement*.

**équivalence** : Une caractéristique. Mesure de la correspondance des données d'une norme ou d'un format à un autre. Des tableaux de concordance ou de conversion sont souvent utilisés à cette fin. L'équivalence entre les codes de la CIM-10-CA et les anciens codes de la CIM-9 en est un exemple.

**erreur de mesure** : Une caractéristique. Différence entre la valeur déclarée dans la banque de données et la valeur réelle, mais inconnue, qui aurait dû être déclarée.

**erreur globale** : Évalue le degré de concordance entre les valeurs déclarées dans la banque de données et celles qui auraient dû l'être, c'est-à-dire les valeurs réelles. Cela équivaut à une vérification de la validité dans la terminologie épidémiologique et donne une indication du nombre de fois où un élément de données est codifié correctement.

**erreur-type** : La racine carrée de la variance.

**estimation** : Regroupement de données destiné à produire une valeur qui représente la population de référence et à tirer des conclusions sur cette population.

**étendue** : Attribut désignant le nombre de caractères permis dans un élément de données.

**étiquette ou signal** : Façon de détecter un cas spécial, souvent au moyen de la création d'un élément de données supplémentaire.

**étude de seconde saisie des données** : Étude dans laquelle des spécialistes désignés reprennent le processus de collecte des données habituellement utilisé par les fournisseurs de données. Les résultats obtenus sont alors comparés aux résultats des fournisseurs de données.

**exactitude** : Une dimension. Mesure dans laquelle l'information provenant ou découlant d'une banque de données reflète la réalité qu'elle doit mesurer.

**facilité d'utilisation** : Une dimension. Mesure de la facilité avec laquelle on peut interpréter et comprendre les données d'une banque de données et y accéder. Désigne aussi la connaissance qu'ont les utilisateurs potentiels de l'existence de la banque de données et du fait que les données sont dans un format convivial et rapidement accessible.

**fardeau de réponse** : Mesure de la difficulté qu'ont les fournisseurs de données à fournir de l'information. Dans de nombreux cas, le fardeau de réponse désigne le temps nécessaire aux fournisseurs pour rassembler et saisir l'information.

**fiabilité** : Voir *cohérence*.

**fiabilité inter-évaluateur** : Taux de divergence obtenu lorsque deux experts en seconde saisie codifient le même élément pour déterminer si leurs réponses concordent.

**fiabilité intra-évaluateur** : Taux de divergence obtenu lorsqu'un expert en seconde saisie doit codifier deux fois le même élément pour déterminer s'il consigne l'information de manière uniforme.

**Fichier de conversion des codes postaux** : Fichier liant les codes postaux aux secteurs de dénombrement.

**formulaire de soumission de données** : Modèle électronique ou sur papier qui indique le format des éléments de données à soumettre.

**fournisseur de données** : Personne ou organisme qui fournit des données à une banque de données. Les fournisseurs de données se chargent habituellement de la saisie des données.

**groupe d'experts** : Groupe de spécialistes en bases de données, tant à l'intérieur qu'à l'extérieur de l'organisme, chargé de donner des conseils sur la mise à jour et l'évolution d'une base de données.

**groupement** : Façon de traiter des variables semblables en les disposant en catégories conçues spécialement pour elles.

**Health Level Seven (HL7)** : Normes pour l'échange, la gestion et l'intégration de données appuyant les soins cliniques et la gestion, la prestation et l'évaluation des services de santé.

**historique des révisions** : Description des modifications ou des corrections apportées à des données présentées dans le passé (données rétrospectives).

**imputation** : Processus de détermination et d'assignation de valeurs de remplacement aux données inexactes ou manquantes qui ne devraient pas l'être lors de la vérification.

**imputation manuelle** : Processus de détermination et d'assignation de valeurs de remplacement aux données inexactes ou manquantes qui ne devraient pas l'être par un processus manuel plutôt qu'automatisé.

**inconnu** : Une cote. Signifie qu'il est impossible de préciser si les exigences du critère ont été satisfaites, ni à quel niveau il est possible de les mesurer.

**interprétabilité** : Une caractéristique. Facilité avec laquelle l'utilisateur peut comprendre les données.

**ISO** : Organisation internationale de normalisation.

**lignes directrices de révision** : Lignes directrices propres à une base de données qui permettent de déterminer s'il est nécessaire ou non de diffuser des données révisées. Les lignes directrices devraient indiquer plus particulièrement à quel moment l'incidence d'erreurs nouvellement découvertes ou de mises à jour serait suffisamment importante pour justifier la diffusion d'un sous-ensemble de données révisées.

**lissage** : Techniques statistiques réduisant les irrégularités (fluctuations aléatoires) dans des séries chronologiques. Elles fournissent un point de vue plus clair du véritable comportement sous-jacent des séries chronologiques.

**manquante de façon aléatoire** : Données aléatoirement absentes d'une banque de données (leur absence n'est pas liée à un autre élément de données).

**méthodes** : Procédures statistiques utilisées dans le traitement et l'analyse des données ou, plus généralement, processus utilisé pour exploiter une base de données.

**microdonnées** : Données servant à l'analyse et à la création de rapports.

**minimal** : Une cote. Signifie que le niveau mesuré est bas.

**mise à jour de la base de sondage** : Ensemble de procédures servant à ajouter toute nouvelle unité de la population de référence dans la base de sondage, de même qu'à retirer toute unité ne faisant plus partie de la population de référence.

**Modèle conceptuel de données sur la santé (MCDS)** : Le MCDS, produit du Partenariat pour les normes d'information sur la santé de l'ICIS, est un modèle généralisé des principaux sujets du domaine des soins de santé dans une perspective canadienne. Tout en fournissant un exemple de l'étendue de l'information qui s'applique aux soins de santé, il indique également comment ces principaux sujets sont liés entre eux.

**modéré** : Une cote. Signifie que le niveau mesuré est moyen.

**non applicable** : Une cote. Signifie que les exigences du critère ne peuvent pas être satisfaites.

**non respecté** : Une cote. Signifie que les exigences du critère n'ont pas été satisfaites.

**non-réponse** : Absence de données pour certaines unités de la base de sondage. Voir aussi *non-réponse totale* et *non-réponse partielle*.

**non-réponse partielle** (ou non-réponse par question) : Une caractéristique. Inclut les éléments de données d'un enregistrement reçu qui sont vides, mais qui ne devraient pas l'être.

**non-réponse totale** : Une caractéristique. Désigne les unités ou les enregistrements entiers appartenant à la base de sondage et *pour lesquels aucune information n'a été présentée*. On la confond souvent avec le sous-dénombrement, c'est-à-dire une information manquante pour des unités qui ne figurent pas dans la base de sondage.

**normalisation** : Une caractéristique. Évaluation du niveau auquel des groupements communs peuvent être obtenus à partir des données.

**normes du Dictionnaire de données** : Une caractéristique. Voir *Dictionnaire de données de l'ICIS*.

**opérationnaliser** : Convertir en une forme utilisable.

**outil d'évaluation** : Composante principale du cadre de la qualité des données. Il se compose de 61 critères utilisés pour définir les aspects problématiques de la qualité des données et évaluer les limites et les forces de la banque de données.

**période de référence** : Période associée aux données en question. Le début de la période de référence correspond à la première date à laquelle les données se rapportent, et la fin de la période de référence, à la dernière date à laquelle les données se rapportent.

**pertinence** : Une dimension. Mesure dans laquelle une banque de données répond aux besoins actuels et futurs des utilisateurs. La pertinence évalue si les données disponibles répondent aux questions les plus importantes des utilisateurs.

**population cible** : Population pour laquelle on désire obtenir de l'information dans une étude statistique. Dans bien des cas, l'information relative à la population cible complète n'est pas disponible. Par exemple, une population cible peut être tous les hôpitaux au Canada ayant au moins un lit de soins de courte durée.

**population de référence** : Population disponible à partir de laquelle on fait des énoncés dans une étude statistique. Par exemple, la population de référence peut être tous les hôpitaux financés par l'État, sauf ceux du système carcéral, qui possèdent au moins un lit de soins de courte durée, dans l'ensemble des provinces et des territoires, et qui étaient ouverts le 1<sup>er</sup> janvier de l'année de référence. La population de référence devrait être aussi proche que possible de la population cible dans toutes les études.

**procédures de mise à jour de la base de sondage** : Pratiques ou procédures utilisées pour mettre à jour la base de sondage.

**programme** : Ensemble d'instructions électroniques pour le traitement des données.

**rapport d'évaluation de la qualité des données** : Rapport interne de l'ICIS qui résume les résultats de l'évaluation de la qualité des données.

**rapport de vérification** : Rapport indiquant si les enregistrements ont réussi ou échoué le processus de vérification de données, et les raisons de l'échec.

**règles de vérification** : Règles servant à déceler les valeurs inexactes ou manquantes dans une banque de données.

**respecté** : Une cote. Signifie que les exigences du critère ont été satisfaites.

**révision** : Modification de données, ou d'estimations tirées de données, une fois que les données sont du domaine public.

**saisie** : Voir *saisie des données*.

**saisie des données** : Entrée des données dans un format utilisable. La saisie des données peut être manuelle ou électronique.

**section** : Groupe à l'ICIS qui travaille à la production d'une base de données ou d'un produit à livrer.

**significatif** : Une cote. Indique un niveau de mesure très élevé. Désigne également un aspect important ou statistiquement significatif.

**signification statistique** : Probabilité que les résultats ne sont pas le fruit du hasard.

**sous-dénombrement** : Absence d'une unité obligatoire dans la base de sondage.

**statistique kappa** : Aussi appelée « coefficient kappa ». Le coefficient kappa mesure la concordance par paire auprès d'un ensemble de codificateurs qui font des jugements et corrige la concordance aléatoire prévue. Le coefficient se calcule ainsi :

$$k = (\text{nombre de concordances observées} - \text{nombre de concordances prévues}) / (\text{nombre total de paires de répondants} - \text{nombre de concordances prévues}).$$
 L'une des méthodes de SAS (PROC FREQ avec /option AGREE) permet aussi de calculer la statistique kappa.

**statistiques agrégées** : S'entend aussi bien des statistiques portant sur un vaste regroupement (ou agrégat) de données, comme des estimations provinciales, que des statistiques utilisées pour résumer (ou agréger) des données, comme la moyenne ou la médiane.

**surdénombrement** : Désigne les cas pour lesquels des unités ne faisant pas partie de la population de référence sont incluses dans la base de sondage, des enregistrements figurent en double dans la base de données ou des enregistrements hors du champ d'observation sont inclus. Voir aussi *enregistrements hors du champ d'observation*.

**tableau de concordance** : Tableau de correspondance pluriunivoque (plusieurs à un ou un à plusieurs) servant à convertir un format de données à un autre. Voir aussi *tableau de conversion*.

**tableau de conversion** : Tableau de correspondance biunivoque (un pour un) servant à convertir un format de données à un autre. Voir aussi *tableau de concordance*.

**tableaux de contrôle** : Voir *tableaux standard*.

**tableaux standard** : Ensemble de tableaux produits à chaque cycle de données et pouvant servir à vérifier les résultats au fil du temps.

**taux de non-réponse partielle** : Pourcentage des éléments de données pour lesquels des données ont été soumises par rapport au nombre d'enregistrements qui auraient dû renfermer des éléments de données.

**taux de réponse des établissements** : Pourcentage des établissements qui ont soumis des données par rapport à tous les établissements de la base de sondage.

**taux de réponse partielle** : Taux d'éléments de données (valeurs) manquants par comparaison avec le nombre d'enregistrements qui ont *de fait* été soumis, et non pas le nombre qui *auraient dû* être soumis. Pour obtenir le taux, multiplier par 100 le nombre d'éléments de données pour lesquels des données ont été déclarées et diviser par le nombre d'enregistrements déclarés qui auraient dû renfermer des éléments de données.

**taux de réponse totale** : Taux calculé à partir du nombre d'unités qui ont soumis des données à une banque de données, multiplié par 100, puis divisé par le nombre d'unités de la base de sondage.

**taux de sous-dénombrement** : Taux calculé à partir du nombre d'unités absentes de la base de sondage mais présentes dans la population de référence, multiplié par 100 et divisé par le nombre d'unités dans la population de référence.

**taux de surdénombrement** : Taux calculé à partir du nombre d'unités présentes dans la base de sondage mais absentes de la population de référence, multiplié par 100 et divisé par le nombre d'unités dans la population de référence.

**traitement** : Application de programmes ou d'une séquence de procédures visant à finaliser une base de données, par exemple pour la production d'estimations, de totaux ou de fréquences, ou encore pour la mise à l'essai des erreurs.

**type de données** : Format d'un élément de données, par exemple format numérique, caractère ou date.

**utilisation** : Mesure de l'utilisation des données d'une banque de données.

**valeur** : Une caractéristique. Mesure de la contribution d'une banque de données aux connaissances sur la santé de la population ou le système de santé et de son utilisation.

**variable non subjective** : Élément de données dont la valeur n'est pas facilement influencée par des croyances ou des sentiments personnels, comme la date de naissance.

**variable subjective** : Élément de données dont la valeur est facilement influencée par des croyances ou des sentiments personnels, comme le degré d'invalidité.

**variance** : Mesure de la variabilité des estimations obtenues à partir du prélèvement de tous les échantillons possibles de la population de référence.

**variations temporelles** : Changements au fil du temps.

**vérification** : Processus permettant de déceler les données manquantes, inexactes ou invalides.

**vérifications de la cohérence** : Combinaison de contrôles effectués sur des éléments de données dans le but d'en vérifier la cohérence (p. ex. un homme qui subit une césarienne est un cas d'incohérence).

**vérifications de la validité** : Contrôles servant à assurer l'exactitude du format des éléments de données et le caractère approprié du taux de réponse. Les vérifications de la validité peuvent consister à comparer la réponse à une liste de réponses acceptables.



---

## Bibliographie

BRACKSTONE, Gordon. « Managing Data Quality in a Statistical Agency », *Survey Methodology* [« La gestion de la qualité des données dans un bureau de statistique », *Techniques d'enquête*], vol. 25, n° 2 (décembre 1999), p. 139-149.

STATISTIQUE CANADA. *Policy on Informing Users of Data Quality and Methodology (Approved March 31, 2000)* [Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie (approuvée le 31 mars 2000)]. Internet :  
<[http://www.statcan.gc.ca/about-aperçu/policy-politique/info\\_user-usager-eng.htm](http://www.statcan.gc.ca/about-aperçu/policy-politique/info_user-usager-eng.htm)> .

STATISTIQUE CANADA. *Statistics Canada Quality Guidelines* [Statistique Canada: lignes directrices concernant la qualité] (cinquième édition), Ottawa (Ont.), Statistique Canada, 2009, n° 12-539-XWE au catalogue.

STATISTIQUE CANADA. *Statistics Canada's Quality Assurance Framework* [Le cadre d'assurance de la qualité de Statistique Canada], Ottawa (Ont.), Statistique Canada, 2002, n° 12-586-XIE au catalogue.

U.S. CENSUS BUREAU. *Definition of Data Quality*, version 1.3, juin 2006. Internet :  
<[http://www.census.gov/quality/P01-0\\_v1.3\\_Definition\\_of\\_Quality.pdf](http://www.census.gov/quality/P01-0_v1.3_Definition_of_Quality.pdf)> .

