



Recommandations sur le nettoyage minimal des données pour l'élaboration de rapports sur la surveillance en prévention et contrôle des infections en soins de courte durée : une solution pour la règle « garbage in, garbage out »

Kathryn Bush^{1*}, Joelle Cayen², Christine Blaser^{3,4,5}, Blanda Chow⁶, Jennifer Ellison⁶, Jennifer Happe^{6,7}, Caroline Quach^{8,9}, Christian Tsang⁶, Olivia Varsaneux², Kristen Versluys¹, Victoria Williams¹⁰, Robyn Mitchell²

Résumé

Contexte : La surveillance des résultats est une composante importante des programmes de prévention et contrôle des infections (PCI), qui permet d'orienter les décisions en matière de soins de santé. Il est essentiel que les données communiquées soient de la plus haute qualité. Il est donc important d'examiner l'exhaustivité, l'exactitude et le caractère opportun des données afin de réduire les incohérences. Toutefois, de nombreux professionnels de PCI n'ont pas été formés aux activités de nettoyage ou de validation de la qualité des données.

Méthodes : Des épidémiologistes experts de tout le Canada ont créé un guide des pratiques exemplaires pour les activités de validation de la qualité des données afin de fournir suffisamment de détails pour améliorer cette activité importante pour la sécurité des patients. La plupart de ces activités sont de simples contrôles visant à vérifier l'exactitude des données sans nécessiter d'examen supplémentaire du dossier du patient ou de couplage avec d'autres ensembles de données.

Résultats : Sur la base d'un consensus entre les experts en surveillance des différentes administrations, des recommandations globales sur la qualité des données dans les programmes de surveillance en PCI ont été élaborées pour améliorer l'exhaustivité (22 %), la précision (68 %) et le caractère opportun (10 %) des données.

Conclusion : La liste des activités de validation de la qualité des données peut être utilisée dans le cadre des activités de surveillance en PCI au Canada afin de soutenir ou d'améliorer les activités de validation de la qualité des données de surveillance existantes pour les programmes de PCI.

Citation proposée : Bush K, Cayen J, Blaser C, Chow B, Ellison J, Happe J, Quach C, Tsang C, Varsaneux O, Versluys K, Williams V, Mitchell R. Recommandations sur le nettoyage minimal des données pour l'élaboration de rapports sur la surveillance en prévention et contrôle des infections en soins de courte durée : une solution pour la règle « garbage in, garbage out ». *Relevé des maladies transmissibles au Canada* 2025;51(9):374–83. <https://doi.org/10.14745/ccdr.v51i09a03f>

Mots-clés : qualité des données, nettoyage des données, surveillance de PCI, exhaustivité, précision, caractère opportun, recommandations

Cette oeuvre est mise à la disposition selon les termes de la licence internationale Creative Commons Attribution 4.0



Affiliations

¹ Interior Health Region, BC

² Programme canadien de surveillance des infections nosocomiales, Agence de la santé publique du Canada, Ottawa, ON

³ Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal, Montréal, QC

⁴ Centre de formation continue, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, QC

⁵ Département de médecine sociale et préventive, École de santé publique, Université de Montréal, Montréal, QC

⁶ Services de santé Alberta, Edmonton, AB

⁷ Groupe d'intérêt sur la surveillance et l'épidémiologie appliquée (GISEA), PCI Canada

⁸ Département de microbiologie, infectiologie et immunologie, Université de Montréal, Montréal, QC

⁹ Centre hospitalier universitaire Sainte-Justine, Montréal, QC

¹⁰ Hôpital Sunnybrook, Toronto, ON

*Correspondance :

kathryn.bush@interiorhealth.ca



Introduction

Les programmes de surveillance en prévention et contrôle des infections (PCI) guident d'importantes décisions en matière de soins de santé, notamment en ce qui concerne l'affectation des ressources aux programmes de PCI et l'élaboration et l'évaluation des mesures de PCI. Les données de surveillance servent à l'évaluation des interventions, à la détection des éclosions afin de prévenir la transmission, au suivi des tendances en matière de préparation aux infections saisonnières et émergentes, et à l'affectation des ressources, y compris le personnel et les protocoles de nettoyage. Les données de surveillance permettent également d'étayer les lignes directrices et les politiques (1). Il est essentiel que les données communiquées soient de la plus haute qualité, et il est donc important de déterminer quels éléments des données de PCI devraient minimalement être examinés et ainsi déterminer les activités de validation de la qualité des données uniformes entre les différents jeux de données, les institutions et les administrations. La qualité des données englobe différents domaines, mais ceux de l'exhaustivité, de l'exactitude et du caractère opportun sont les plus importants (2,3). Les données sont complètes lorsque tous les patients admissibles sont inclus dans les cas de surveillance et que toutes les variables du formulaire de saisie des données de surveillance sont remplies. Les données sont exactes lorsque les cas correspondent à la définition de cas du protocole et que les décisions de classification des données sont correctes. Le caractère opportun des données correspond à leur disponibilité et diffusion lorsque les résultats sont requis (2,3).

Le Programme canadien de surveillance des infections nosocomiales (PCSIN) est une collaboration entre l'Association pour la microbiologie médicale et l'infectiologie Canada (AMMI) et l'Agence de la santé publique du Canada (l'Agence) pour effectuer une surveillance standardisée de la résistance aux antimicrobiens et des organismes résistants aux antimicrobiens dans les établissements de soins de courte durée sentinelles du Canada (1). Le groupe de travail sur la qualité des données du PCSIN se compose d'experts canadiens de la surveillance qui possèdent collectivement une expérience en matière de prévention et de contrôle des infections, de surveillance, d'épidémiologie et d'expertise clinique. Il s'agit de membres du personnel du PCSIN ou de représentants des sites du réseau de surveillance canadien. Le groupe a mené des activités de validation de la qualité des données pour soutenir la surveillance du PCSIN depuis 2005.

L'objectif de ce projet consistait à décrire des activités de validation de la qualité des données suffisamment détaillées pour soutenir les activités de validation de la qualité des données de surveillance existantes en PCI avant l'analyse et l'élaboration de rapports par les équipes canadiennes de PCI.

Méthodes

Ce projet du PCSIN s'est appuyé sur les travaux initiaux des analystes du programme de PCI d'Alberta Health Services (AHS). La liste a été initialement établie sur la base des activités relatives aux données que les analystes en PCI de l'AHS ont entreprises pour l'examen et la validation de la qualité des données de chaque initiative de surveillance du PCSIN, y compris les organismes résistants aux antimicrobiens, l'infection à *Clostridioides difficile* et les infections des sites opératoires, les bactériémies et les virus respiratoires. Les activités se sont principalement concentrées sur les éléments de données qui exigent que la personne qui collecte les données interprète les événements cliniques ou qui sont sujets à des erreurs de saisie (p. ex., les champs de date).

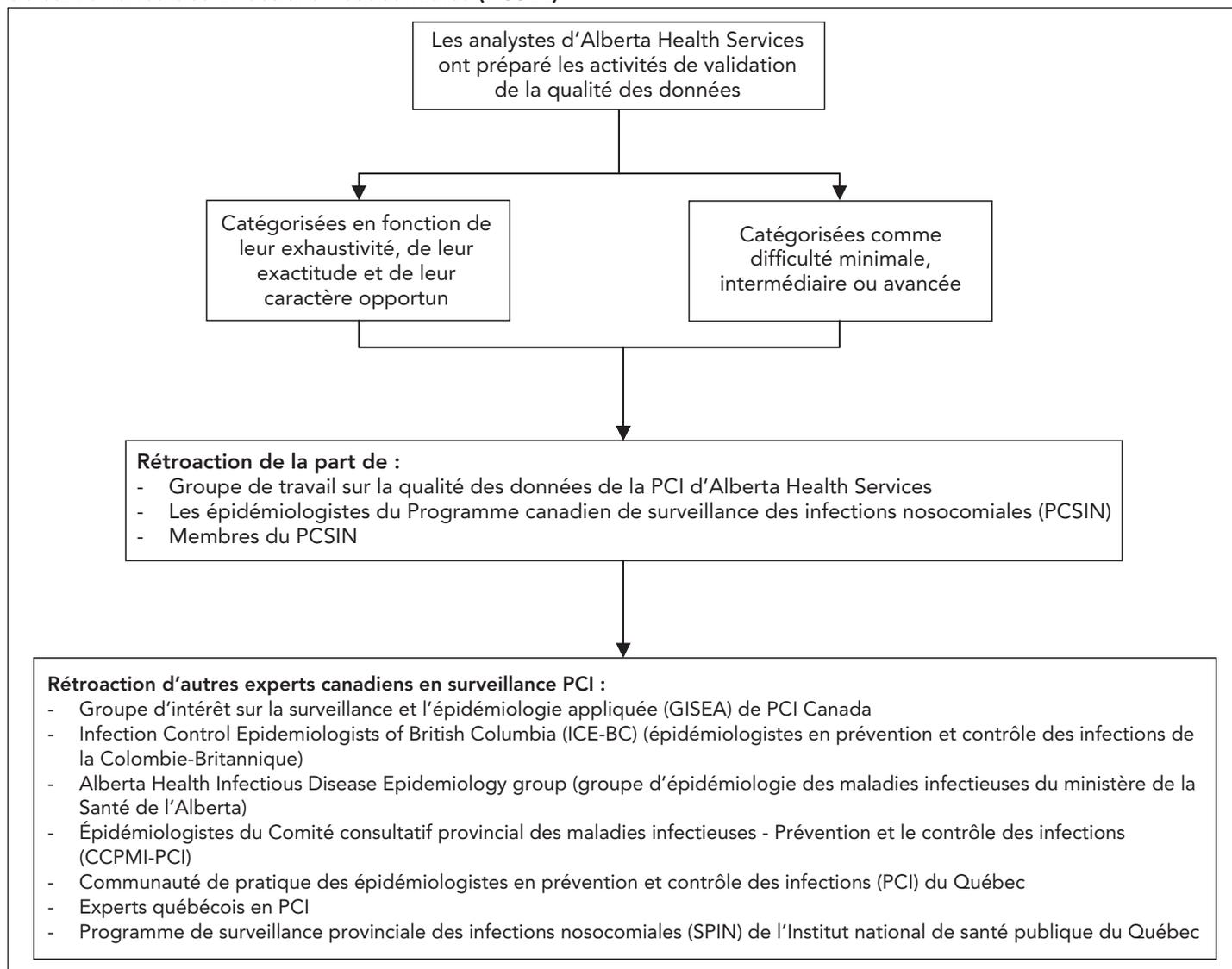
À la suite de ce développement initial, le groupe de travail sur la qualité des données du PCSIN a affiné la liste et a ensuite demandé l'avis et l'approbation d'experts en surveillance partout au Canada (**figure 1**).

Les activités de validation de la qualité des données concernent à la fois les éléments de données que les processus de traitement des données. Un élément de données est la plus petite unité de données et représente un cas, un événement ou un individu (p. ex., le résultat d'un test de laboratoire, le sexe à la naissance, l'âge, etc.). Un processus de traitement des données fait référence aux étapes de la collecte, de la gestion et de l'analyse des éléments de données, transformant les données brutes en renseignements concrets.

Les activités particulières de validation de la qualité des données ont été classées en trois catégories : exhaustivité, exactitude et caractère opportun, sur la base des définitions de ces domaines de qualité des données (2,3). Le niveau de difficulté reflète la capacité à examiner les données sans sources de données supplémentaires ou sans compétences avancées en matière de gestion des données. Les activités d'un niveau de difficulté minimale sont celles qui ne nécessitent qu'un examen des données sans référence à d'autres sources de données. Les activités de difficulté intermédiaire demandent l'examen des données et du dossier du patient. Les activités de difficulté avancée sont basées sur l'établissement de liens avec de grands ensembles de données administratives (p. ex., la base de données sur les congés) ou des activités nécessitant l'examen d'un médecin expert.



Figure 1 : Élaboration de la liste des activités de validation liées à la qualité des données du Programme canadien de surveillance des infections nosocomiales (PCSIN)



Abréviations : AHS, Alberta Health Services; CCPMI-PCI, Comité consultatif provincial des maladies infectieuses - Prévention et contrôle des infections; GISEA, Groupe d'intérêt sur la surveillance et l'épidémiologie appliquée; PCI, Prévention et contrôle des infections; PCSIN, Programme canadien de surveillance des infections nosocomiales; SPIN, Surveillance provinciale des infections nosocomiales

Résultats

Après la validation des domaines de qualité des données et des niveaux de difficulté, l'analyse a montré que les activités relatives à la validation de la qualité des données concernaient principalement l'exactitude des données (68 %), avec moins d'éléments relatifs à l'exhaustivité (22 %) et au caractère opportun (10 %). La plupart des activités liées à la qualité des données ont été classées comme des activités de difficulté minimale (71 %), avec quatre activités de difficulté intermédiaire (13 %) et cinq activités de difficulté avancée (16 %) (tableau 1).

Les rétroactions des différents experts canadiens en épidémiologie ont été recueillies sous forme d'entrevues, ce qui

Tableau 1 : Domaines de qualité des données par niveau de difficulté pour les activités liées à la qualité des données

Domaines de la qualité des données	Niveau de difficulté de la validation			Total
	Difficulté minimale ^a	Difficulté intermédiaire ^b	Difficulté avancée ^c	
Intégralité	3	2	2	7 (2 %)
Exactitude	17	2	2	21 (68 %)
Caractère opportun	2	0	1	3 (10 %)
Total	22 (71 %)	4 (13 %)	5 (16 %)	31

^a Activités de difficulté minimale : celles qui ne nécessitent qu'un examen des données sans référence à une autre source de données

^b Activités de difficulté intermédiaire : celles qui nécessitent l'examen des données et du dossier du patient

^c Activités de difficulté avancée : celles qui nécessitent des liens avec de grands ensembles de données administratives (p. ex., la base de données sur les congés) ou celles qui nécessitent l'examen par un médecin expert



a mené à des mises à jour de la liste des activités de validation de la qualité des données. Les commentaires des experts étaient positifs, avec un consensus concernant la liste proposée d'activités de validation de la qualité des données (**tableau 2**). Par exemple, un expert a indiqué qu'il incluait ces activités dans sa pratique, mais qu'il était ouvert aux suggestions concernant d'autres activités de validation qu'il n'avait pas envisagées.

La liste des activités a ensuite été approuvée par tous les groupes contactés. D'autres mesures ont été effectuées suite à la discussion avec le Groupe d'intérêt sur la surveillance et l'épidémiologie appliquée (GISEA) de PCI Canada, y compris les discussions avec la direction de PCI Canada, qui ont abouti à l'approbation de la liste par le conseil d'administration de PCI Canada.

Tableau 2 : Liste des activités de validation de la qualité des données du Programme canadien de surveillance des infections nosocomiales

Domaines de la qualité des données	Élément de données, processus de données	Niveau de difficulté	Processus	Renseignements complémentaires/ activité/mesure
Exhaustivité – tous les cas admissibles sont inclus; toutes les variables de données sont remplies.	Nom du patient, sexe, numéro(s) d'identification du dossier de santé	Difficulté minimale	Vérifier l'exactitude des données saisies par rapport à la source (dossier clinique).	Intégrer des options pour établir des liens avec d'autres noms pour la même personne.
	Données manquantes dans les variables	Difficulté minimale	Vérifier dans le système clinique pour s'assurer que les données saisies sont complètes.	Envisager des règles de validation des données dans le système de surveillance pour les éléments de saisie obligatoires. Examen périodique des éléments de données manquants afin de déterminer si les données doivent continuer à être recueillies (p. ex., les facteurs de risque d'infection par des organismes producteurs de carbapénémase [OPC] ou d'infection respiratoire virale [IRV]). Fournir un rapport aux responsables de la surveillance sur les variables de données à faible taux de réponse.
	Cas manquants	Difficulté avancée	Confirmer que tous les échantillons positifs et admissibles sont pris en compte pour la surveillance.	Confirmer que toutes les hémocultures prélevées lors de la préadmission (y compris aux urgences) lorsque les patients sont directement transférés vers des soins hospitaliers sont incluses car les hémocultures peuvent ne pas être prélevées à nouveau lors de l'admission du patient. Envisager l'examen des données de laboratoire pour confirmer que tous les cas de surveillance ont été pris en compte. Envisager l'examen des données relatives au dénominateur des interventions chirurgicales pour confirmer que tous les patients inclus remplissent les conditions d'éligibilité à la surveillance.
	Saisie des variables de résultats (p. ex., décès attribuable, admission attribuable aux soins intensifs, colectomie attribuable, autres événements indésirables attribuables)	Difficulté avancée	Ces événements indésirables sont suivis pendant 30 jours après l'événement de surveillance. Pour garantir une saisie complète des cas, des couplages de données peuvent être effectués avec les données administratives de congé afin de confirmer que tous les patients présentant un événement indésirable éligible ont été saisis.	Les données sur les congés des patients (DCP) contiennent des renseignements sur la raison du congé, le diagnostic et l'admission en unité de soins intensifs (USI). Les données relatives à l'admission, au congé et au transfert (ACT) contiennent des renseignements sur tous les mouvements du patient au cours d'une admission.
	Liens entre le dénominateur et le numérateur	Difficulté intermédiaire	Cas inclus qui n'est pas présent dans le dénominateur.	Examen par l'analyste, puis suppression du dossier.



Tableau 2 : Liste des activités de validation de la qualité des données du Programme canadien de surveillance des infections nosocomiales (suite)

Domaines de la qualité des données	Élément de données, processus de données	Niveau de difficulté	Processus	Renseignements complémentaires/ activité/mesure
Exhaustivité – tous les cas admissibles sont inclus; toutes les variables de données sont remplies. (suite)	Dossiers « complets »	Difficulté minimale	Une fois les contrôles de qualité des données terminés, marquez le dossier comme finalisé et n'autorisez pas d'autres modifications, sauf si elles sont documentées.	Les dossiers « complets » peuvent être retirés des extractions de données de routine à des fins de nettoyage ou d'examen.
Précision – le cas répond à la définition de surveillance; les décisions de classification sont correctes.	Doublons	Difficulté minimale	Vérifier qu'il n'y a pas de doublons dans les noms et les dossiers des patients.	Supprimer les doublons, autoriser seulement un cas de surveillance par période de protocole.
	Les cas de surveillance répondent aux définitions de cas du protocole et aux critères d'éligibilité (synonymes : cas incident, cas primaire, cas de surveillance)	Difficulté minimale	Vérifier l'admission dans un centre de soins de longue durée au moment de la détection.	Créer des décisions d'interprétation de protocole pour les patients hospitalisés d'urgence, les évaluations de préadmission, les soins urgents avec transfert direct vers les soins de longue durée.
	Décision de classification des cas	Difficulté minimale	Vérifier le temps écoulé entre l'admission et la date de mise en culture.	Calculer le « délai entre » pour confirmer la décision de classification avec le protocole (p. ex., acquis dans la communauté ou dans les soins de santé) ou l'admissibilité du cas.
	Début de l'infection	Difficulté minimale	Vérifier que la date d'apparition de l'infection correspond à la définition du cas de surveillance.	Utiliser des définitions d'infection de référence pour la comparabilité avec d'autres systèmes de surveillance (p. ex., PCSIN, National Healthcare Surveillance Network [NHSN]) pour les décisions en matière d'infection et la surveillance des bactériémies et du site opératoire. Examiner le dossier clinique pour s'assurer de l'accord avec la décision du professionnel en prévention et contrôle des infections (PCI).
	Délai entre l'admission et la date de prélèvement	Difficulté minimale	Pour la surveillance des bactériémies associées aux cathéters centraux (BACC), vérifier que le cathéter central était en place pendant la durée minimale avant que la culture ne soit positive.	Calculer le « délai entre » la date d'insertion et la date de prélèvement et entre la date d'admission à l'USI et la date de prélèvement pour déterminer si le cas répond à la définition du protocole.
	Cas de surveillance valide pour les entrées de données multiples	Difficulté minimale	Vérifier le délai entre les cas de surveillance des infections nosocomiales (4).	Confirmer le nouveau cas avec la définition du protocole (4).
	La date de naissance correspond à la date d'admission à l'hôpital	Difficulté minimale	Indique une erreur de saisie pour les adultes.	Modifier le dossier pour corriger les erreurs. Remarque : peut être correct pour les nouveau-nés.
	La date de naissance correspond à la date de prélèvement du laboratoire	Difficulté minimale	Indique une erreur de saisie des données.	Remarque : peut être correct pour les nouveau-nés.
	La date de naissance correspond à la date d'apparition de l'infection	Difficulté minimale	Indique une erreur de saisie des données.	Remarque : peut être correct pour les nouveau-nés.
La date d'admission à l'hôpital correspond à la date de prélèvement	Difficulté minimale	Confirmer s'il s'agit d'une infection nosocomiale.	Confirmer l'échéancier de la définition de la classification des cas.	

Tableau 2 : Liste des activités de validation de la qualité des données du Programme canadien de surveillance des infections nosocomiales (*suite*)

Domaines de la qualité des données	Élément de données, processus de données	Niveau de difficulté	Processus	Renseignements complémentaires/ activité/mesure
Précision – le cas répond à la définition de surveillance; les décisions de classification sont correctes. (<i>suite</i>)	La date de l'infection correspond à la date de l'intervention	Difficulté minimale	S'applique à la surveillance des infections du site opératoire.	Un cas d'infection du site opératoire ne peut pas survenir le même jour que la première intervention chirurgicale.
	Erreurs de formatage dans les champs de date	Difficulté minimale	Erreur de saisie des données, par exemple, mois/année.	Utiliser le format de date aaaa/mmm/jj.
	Nom de l'organisme (OPC, bactériémie, BACC, infection du site opératoire, IRV)	Difficulté minimale	Confirmer et corriger les noms des agents pathogènes et faire la distinction entre les différentes souches d'un même agent pathogène, déterminer l'admissibilité des cas pour les commensaux communs et traiter les données relatives à plusieurs agents pathogènes au sein d'un même cas.	Comparer les données au dossier médical clinique et aux résultats du laboratoire. Utiliser le tableau des « commensaux communs » du NHSN comme modèle.
	Site de culture pour la surveillance des organismes résistants aux antimicrobiens (ORA)	Difficulté minimale	Confirmer que le site de prélèvement clinique est saisi.	Si la définition de surveillance exige la présence d'une infection, les sites de dépistage des organismes résistants aux antimicrobiens (p. ex., les échantillons de dépistage nasal) ne sont pas autorisés pour les cas de surveillance.
	Réponses « Autres »	Difficulté minimale	Confirmer que « Autres » n'est pas sélectionné si la réponse valide est incluse dans une liste de contrôle ou un menu déroulant.	Examiner toutes les entrées de texte libre afin d'en minimiser l'utilisation.
	Erreurs de formatage dans les variables de données textuelles	Difficulté minimale	Exemples : traits d'union inclus ou non inclus dans les numéros d'accès, variabilité du nom du site ou de l'unité (acronyme, fautes d'orthographe).	Examiner la qualité des données en fonction des normes du système. Envisager des champs déroulants dans le système de saisie des données pour plus de cohérence.
	Date de décès valide	Difficulté minimale	Ne pas inclure les décès survenus après la période couverte par le protocole.	La plupart des protocoles demandent un résultat 30 jours après la date de l'enregistrement de la surveillance. Ne pas inclure les décès ou autres résultats supérieurs à 30 jours.
	Données de laboratoire : plusieurs échantillons prélevés	Difficulté intermédiaire	Examen des dossiers de laboratoire et/ou de surveillance par rapport au dossier clinique : si plusieurs échantillons ont été prélevés dans différentes unités, les échantillons positifs des patients hospitalisés sont sélectionnés en fonction de la population de patients de chaque protocole.	Créer des indications supplémentaires pour la saisie des données à l'intention des utilisateurs afin de permettre de futurs couplages de données de laboratoire à des fins de recherche de cas – p. ex., si les échantillons ont des niveaux similaires de pertinence clinique (p. ex., urine et plaie), l'échantillon prélevé en premier est sélectionné. S'il existe plusieurs numéros d'accès pour le même microorganisme provenant de plusieurs sites d'échantillonnage, l'échantillon le plus pertinent d'un point de vue clinique est sélectionné. Si plusieurs spécimens sont prélevés en même temps, le spécimen rapporté en premier est sélectionné.



Tableau 2 : Liste des activités de validation de la qualité des données du Programme canadien de surveillance des infections nosocomiales (suite)

Domaines de la qualité des données	Élément de données, processus de données	Niveau de difficulté	Processus	Renseignements complémentaires/ activité/mesure
Précision – le cas répond à la définition de surveillance; les décisions de classification sont correctes. (suite)	Renseignements sur la rencontre : établissement, unité, service, lit	Difficulté intermédiaire	Le cas de surveillance est saisi pour le lieu auquel le cas est attribué plutôt que pour le lieu où le cas est détecté.	Examiner le dossier clinique, envisager une procédure standard pour saisir les cas de patients qui ont été transférés plusieurs fois (c.-à-d., la dernière unité dans laquelle le patient se trouvait par rapport à l'unité dans laquelle le patient se trouvait au moment du prélèvement de l'échantillon).
	Décision relative à l'infection/la colonisation	Difficulté intermédiaire	Renvoyer les cas aux collecteurs de données pour un nouvel examen, afin de confirmer la décision d'infection relative au NHSN.	Activité périodique sur la qualité des données : en particulier pour les cas avec des échantillons d'expectoration, de plaies et d'urine, qui ont une probabilité plus élevée de colonisation par des organismes.
	Symptômes pour la surveillance de l'IRV	Difficulté intermédiaire	Examiner le nom du virus et les symptômes associés pour déterminer l'admissibilité du cas de surveillance.	L'agent pathogène peut déterminer quels symptômes font partie d'un cas de surveillance.
	Statut vaccinal	Difficulté avancée	Ne peut pas indiquer COVID-19 ou d'autres vaccinations avant la disponibilité des vaccins.	Nécessité de suivre les différentes dates de disponibilité des vaccins et leur indication dans chaque province.
Caractère opportun – les données de surveillance sont fournies au moment où elles sont requises.	Temps écoulé entre le résultat de la culture et la saisie des données	Difficulté minimale	Tenir compte des délais de saisie des données afin de fournir des données en temps utile, p. ex., les données saisies dans les 5 jours suivant la date de la culture positive.	Délai entre la date de saisie des données et la date du résultat de la culture – fournir une rétroaction à la PCI et aux responsables, s'efforcer de créer des efficacités pour permettre une saisie rapide des données de surveillance.
	Délai d'examen	Difficulté minimale	Déterminer la fréquence des activités liées à la qualité des données.	Examiner chaque type de contrôle de la qualité des données avec les calendriers de déclaration afin de déterminer la fréquence ou le calendrier de nettoyage des données (quotidien, hebdomadaire, mensuel).
	Délai d'établissement des rapports	Difficulté minimale	Déterminer la fréquence des rapports (quotidienne, hebdomadaire, mensuelle, trimestrielle, annuelle).	Créer des activités et des calendriers relatifs à la qualité des données afin de tenir compte de la fréquence des rapports, y compris le temps nécessaire à l'examen avec la PCI.
	Dénominateurs pour la production de rapports	Difficulté avancée	Déterminer la fréquence des rapports préliminaires et définitifs sur les taux.	Tenir compte de la source des données du dénominateur et de la possibilité d'un processus de recherche de cas en temps opportun pour répondre aux besoins des parties prenantes et à la disponibilité des données. Inclus : jours-présence, admissions, jours-cathéters, procédures chirurgicales.

Abréviations : ACT, admission, congé, transfert; BACC, bactériémies associées aux cathéters centraux; DCP, données sur les congés des patients; IRV, infection respiratoire virale; ISO, infections du site opératoire; NHSN, National Healthcare Surveillance Network; OPC, organismes producteurs de carbapénémase; ORA, organisme résistant aux antibiotiques; PCI, prévention et contrôle des infections; PCSIN, Programme canadien de surveillance des infections nosocomiales; USI, unité de soins intensifs

Discussion

L'objectif de ce projet consistait à fournir des activités de validation de la qualité des données suffisamment détaillées pour soutenir les données de surveillance existantes en PCI, avant l'analyse et la déclaration par les professionnels canadiens en PCI. À notre connaissance, il s'agit de la première liste

détaillée d'activités relatives à la qualité des données pour le personnel chargé de rapporter les résultats de surveillance en PCI au Canada. La collaboration entre les experts canadiens en surveillance de PCI a été importante pour valider ces activités de validation de la qualité des données et pour les communiquer aux programmes de PCI de l'ensemble du pays. D'autres programmes ont recommandé des activités de validation de



la qualité des données, notamment le *National Healthcare Surveillance Safety Network* (NHSN) des *Centers for Disease Control*, l'Institut canadien d'information sur la santé (ICIS) et l'*Australian Commission on Safety and Quality in Health Care* (ACSQHC) (3,5–7). Ces experts recommandent de créer un plan de surveillance global pour comprendre les sources de données, la validation des données (y compris les liens entre les données administratives et les données de laboratoire pour valider la recherche de cas) et d'autres vérifications au niveau du système, mais à l'exception de la ressource australienne, aucun ne propose d'éléments spécifiques.

La surveillance des infections nosocomiales est l'une des principales responsabilités de tout programme de PCI qui considère que la diligence raisonnable dans la création de données complètes, exactes et opportunes est une exigence importante avant la communication des résultats. Une publication récente estime que 45 % du temps d'un professionnel en prévention et contrôle des infections est consacré aux activités de surveillance et que les systèmes manuels peuvent avoir une précision de seulement 62,5 % (étendue : 16 %–87 %) (8). Sur la base de cette précision sous-optimale des données déclarées, nous avons cherché à développer une liste d'activités de validation pour aider à améliorer la qualité des données de surveillance en PCI. Bien que la liste des activités liées à la qualité des données du PCSIN soit conçue pour la surveillance du PCSIN, son intérêt réside dans le fait que tout professionnel en PCI qui recueille et communique des données de surveillance peut utiliser les activités suggérées pour améliorer la qualité des données de son programme.

Bien que cette liste permette de vérifier l'exhaustivité des données, les informations difficiles à recueillir et souvent non déclarées doivent être régulièrement examinées afin de déterminer leur utilité pour les rapports de surveillance. L'identification des éléments de données nécessaires et l'examen des données incomplètes peuvent souvent orienter la saisie des données vers les données essentielles requises pour la production de rapports. Un ensemble de données de base est généralement créé à partir d'un consensus d'experts pour les éléments de données essentielles pour la production de rapports, afin de réduire la charge de la collecte de données et d'améliorer la qualité des données déclarées (9). Les activités de validation de la qualité des données décrites dans cet article impliquent l'utilisation d'autres sources de données pour confirmer que la recherche de cas est équivalente dans l'ensemble du système de surveillance et que tous les cas potentiels sont examinés, même s'ils ont été omis par le collecteur de données initial.

L'exactitude des données commence par la formation et le soutien des personnes chargées de la collecte des données en ce qui concerne les définitions de cas et les critères d'inclusion et d'exclusion telles que décrites dans le protocole. Pour évaluer l'application des définitions du protocole du PCSIN dans les

hôpitaux qui participent à la surveillance du PCSIN, le groupe de travail sur la qualité des données a mené plusieurs études. Ces résultats montrent un taux de réponses correctes de 88 % pour la surveillance des bactériémies (10), de 79 % pour la surveillance de la COVID-19 (11) et de 78 % pour la surveillance des infections à *Clostridioides difficile* (non publié). Des rapports anecdotiques ont indiqué que les sites continuent d'utiliser les questions de l'enquête lors de l'orientation des nouveaux membres de la PCI afin de discuter de la bonne interprétation du protocole du PCSIN.

Dans le domaine du respect des délais, il est nécessaire que chaque réviseur établisse un calendrier de publication des rapports de surveillance, afin que les attentes en matière de contrôle de la fréquence des activités liées à la qualité des données et de respect des délais de saisie des données puissent être définies. Des audits supplémentaires peuvent être programmés pour vérifier ponctuellement la qualité des données par un autre réviseur, à titre d'étape de contrôle de la qualité.

L'ICIS prévoit deux domaines supplémentaires de validation de la qualité des données : la pertinence (les données répondent aux besoins actuels et potentiels des utilisateurs) et l'accessibilité (les résultats de la surveillance sont facilement accessibles et présentés de manière claire et compréhensible) (5). Il s'agit de deux domaines importants qui permettent aux partenaires cliniques de comprendre et d'utiliser les données. Les discussions avec les partenaires cliniques peuvent confirmer que les infections sous surveillance et les données présentées sont exploitables, et approuver les définitions de surveillance afin d'établir les cas cliniquement pertinents.

Dans un souci de généralisation, la liste des activités du PCSIN relatives à la qualité des données est conçue pour différents niveaux de contrôle. Les contrôles minimaux, que tout membre du personnel de la PCI peut effectuer, représentent 71 % (22/31) de l'ensemble des activités, dont 77 % (17/22) sont des contrôles d'exactitude. Cela permet une certaine souplesse dans l'exécution des contrôles de qualité des données pour les systèmes de surveillance qui n'ont pas les ressources nécessaires pour effectuer des contrôles de niveau intermédiaire ou avancé. Un réviseur a reconnu l'utilité des analystes en surveillance en PCI pour aider l'épidémiologiste à réaliser les activités de validation de la qualité des données, et ces activités de pratiques exemplaires peuvent justifier les postes d'analystes dans le cadre d'une analyse de rentabilisation.

Limites

Toute mesure de l'amélioration des données locales grâce à ces activités de validation de la qualité des données dépend de l'état des données consignées à l'origine. Certains sites peuvent avoir mis en place d'autres interventions, telles que des initiatives de formation continue pour aider les collecteurs de données à interpréter plus précisément les protocoles, ce qui aurait une incidence sur l'amélioration globale des données du site en



question. Cependant, la règle fondamentale de l'intégrité des données (« garbage in, garbage out » ou « à données inexactes, résultats erronés ») s'applique à toutes les données. Les données de surveillance en PCI ont l'avantage d'être générées par les professionnels de la prévention et contrôle des infections en tant que collecteurs de données primaires, et ont donc l'avantage, par rapport aux données administratives, de pouvoir fait l'objet d'activités de validation de la qualité globale des données. Ceci évite les problèmes liés aux données administratives, tels que des conclusions trompeuses en raison de l'imprécision globale des données (12).

Les prochaines étapes comprennent la diffusion de la liste des activités liées à la qualité des données et la mise en œuvre d'une formation au sein du réseau du PCSIN et pour les collecteurs de données de surveillance de PCI Canada. Une évaluation de l'utilité, de la pertinence, de l'exhaustivité et de l'efficacité de la liste suivra lorsqu'elle aura été utilisée pendant quelques années, et une version ultérieure inclura toutes les suggestions d'amélioration.

Conclusion

Les activités liées à la qualité des données décrites constituent une liste qui établit un ensemble standardisé d'activités de validation de la qualité des données afin de garantir une communication complète, précise et opportune des données de surveillance en PCI. Les activités ont été examinées par des experts en surveillance en PCI à travers le Canada et validées par le programme de PCI des Alberta Health Services. Elles peuvent être utilisées pour soutenir ou améliorer les activités liées à la qualité des données de surveillance existantes dans les équipes de PCI. L'application systématique de la liste des activités de validation de la qualité des données du PCSIN va soutenir ou améliorer les activités de validation de la qualité des données de surveillance et accroître la confiance des utilisateurs des résultats de la surveillance.

Déclaration des auteurs

K. B. — Conceptualisation, interprétation, rédaction de la version originale

J. C. — Validation, rédaction-révision et édition

K. V. — Validation, rédaction-révision et édition

O. V. — Validation, rédaction-révision et édition

R. M. — Validation, rédaction-révision et édition

C. B. — Validation, rédaction-révision et édition

C. O. — Validation, rédaction-révision et édition

V. W. — Validation, rédaction-révision et édition

J. H. — Validation, rédaction-révision et édition

B. C. — Méthodologie, rédaction-révision et édition

J. E. — Méthodologie, rédaction-révision et édition

C. T. — Enquête, rédaction-révision et édition

Intérêts concurrents

Les auteurs n'ont aucun conflit d'intérêts à divulguer.

Identifiants ORCID

Kathryn Bush — [0000-0001-8297-5298](https://orcid.org/0000-0001-8297-5298)

Christine Blaser — [0000-0001-6565-1742](https://orcid.org/0000-0001-6565-1742)

Jennifer Ellison — [0000-0001-6278-9519](https://orcid.org/0000-0001-6278-9519)

Jennifer Happe — [0009-0007-5407-6552](https://orcid.org/0009-0007-5407-6552)

Caroline Quach — [0000-0002-1170-9475](https://orcid.org/0000-0002-1170-9475)

Christian Tsang — [0000-0002-5797-8523](https://orcid.org/0000-0002-5797-8523)

Olivia Varsaneux — [0000-0003-3249-0721](https://orcid.org/0000-0003-3249-0721)

Kristen Versluys — [0000-0003-4680-3311](https://orcid.org/0000-0003-4680-3311)

Robyn Mitchell — [0000-0002-6241-196X](https://orcid.org/0000-0002-6241-196X)

Remerciements

Les auteurs remercient les épidémiologistes canadiens d'avoir apporté leur expertise et tous les professionnels en prévention et contrôle des infections d'avoir fourni des données de surveillance complètes, précises et opportunes pour assurer la sécurité des patients au Canada.

Financement

Il n'y a pas de financement pour ce travail.

Références

1. Programme canadien de surveillance des infections nosocomiales. Le Programme canadien de surveillance des infections nosocomiales : suivi de la résistance aux antimicrobiens dans les hôpitaux canadiens depuis 1995. Relevé des maladies transmissibles au Canada 2022;48(12):506–11. [DOI](#)
2. Gouvernement du Canada. Orientation sur la qualité des données. Ottawa, ON : Gouvernement du Canada; 2024. <https://www.canada.ca/fr/gouvernement/systeme/gouvernement-numerique/innovations-gouvernementales-numeriques/gestion-information/orientation-qualite-donnees.html>
3. Kilkeny MF, Robinson KM. Data quality: "Garbage in - garbage out". HIM J 2018;47(3):103–5. [DOI PubMed](#)
4. Prévention et contrôle des infections Canada Protocoles et publications du PCSIN. Winnipeg, MB : PCI Canada; 2025. <https://ipac-canada.org/fr/resource-centre/infection-control-resources/cnisp-protocoles-publications/>



5. Australian Commission on Safety and Quality in Health Care. Validating infection surveillance data factsheet. Sydney, NSW: ACSQHC; 2024. https://www.safetyandquality.gov.au/sites/default/files/2024-04/data_validation_guide_factsheet.pdf
6. Institut canadien d'information sur la santé. Cadre de la qualité de l'information de l'ICIS. Ottawa, ON : ICIS; 2024. <https://www.cihi.ca/sites/default/files/document/information-quality-framework-nov-2024-fr.pdf>
7. Center for Disease Control. NHSN 2019 Guidance and Toolkit for Data Quality Checks for Reporting Facilities. Atlanta, GA: CDC; 2019. <https://www.cdc.gov/nhsn/pdfs/validation/2019/2019-nhsn-iv-for-facilities-508.pdf>
8. Garcia R, Barnes S, Boukidjian R, Goss LK, Spencer M, Septimus EJ, Wright MO, Munro S, Reese SM, Fakih MG, Edmiston CE, Levesque M. Recommendations for change in infection prevention programs and practice. *Am J Infect Control* 2022;50(12):1281–95. [DOI PubMed](#)
9. Soucie JM. Public health surveillance and data collection: general principles and impact on hemophilia care. *Hematology* 2012;17(Suppl 1):S144–6. [DOI PubMed](#)
10. Ellison J, Cayen J, Pelude L, Mitchell R, Bush K. Evaluation of the accuracy in the application of the CNISP bloodstream infection surveillance definitions. *Cdn J Inf Control* 2023;38:19–22. [DOI](#)
11. McGill E, Cayen J, Ellison J, Lee, D, Pelude L, Mitchell R, Frenette C, Thampi N, Bush K. An assessment of the validity and reliability of SARS-CoV-2 infection surveillance data from the Canadian Nosocomial Infection Surveillance Program (CNISP). *Cdn J Inf Control* 2023;38:112–6. [DOI](#)
12. Grimes DA. Epidemiologic research using administrative databases: garbage in, garbage out. *Obstet Gynecol* 2010;116(5):1018–9. [DOI PubMed](#)



Canada

Aimeriez-vous
publier un
article dans
le **RMTC**?

Faites-nous
parvenir votre
manuscrit!



Visite : phac-aspc.gc.ca/publicat/ccdr-rmtc/ia-ra-fra.php