

Poids à la naissance et durée de gestation manquants? La solution dans le jumelage des fichiers des naissances et des hospitalisations

MISSING DATA ON BIRTH WEIGHT AND GESTATIONAL AGE? A SOLUTION BY MATCHING BIRTH AND HOSPITALIZATION RECORDS

PESO AL NACIMIENTO Y DURACIÓN DE GESTACIÓN FALTANTES : LA SOLUCIÓN EN EL EMPAREJAMIENTO DE LOS REGISTROS DE LOS NACIMIENTOS Y DE LAS HOSPITALIZACIONES

Sylvie Montreuil, Richard Halley and Shirley Joe

Volume 25, Number 2, Fall 1996

La santé

URI: <https://id.erudit.org/iderudit/010212ar>

DOI: <https://doi.org/10.7202/010212ar>

[See table of contents](#)

Publisher(s)

Association des démographes du Québec

ISSN

0380-1721 (print)

1705-1495 (digital)

[Explore this journal](#)

Cite this note

Montreuil, S., Halley, R. & Joe, S. (1996). Poids à la naissance et durée de gestation manquants? La solution dans le jumelage des fichiers des naissances et des hospitalisations. *Cahiers québécois de démographie*, 25(2), 261–278. <https://doi.org/10.7202/010212ar>

Article abstract

The high percentage of missing data on birth weight and gestational age in Quebec's birth registration file made it difficult to calculate certain health indicators. However, these data were available in the hospitalization files. Matching the two files made it possible to find the same individual's record in each file, fill in the missing data and improve the quality of the birth records. The matching proved successful in that it reduced the percentage of missing values in the birth registration file to 0.1% for birth weight and 0.2% for gestational age in 1993, compared to 1.1% and 3.2% respectively in 1992. The same procedure was used for stillbirth records, with even more dramatic results: the percentages of missing values fell from 10.4% and 7.7% respectively in 1992 to 0.3% in both categories in 1993.

Tous droits réservés © Association des démographes du Québec, 1996

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

Érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

Poids à la naissance et durée de gestation manquants ? La solution dans le jumelage des fichiers des naissances et des hospitalisations

Sylvie MONTREUIL, Richard HALLEY et Shirley JOE *

Les événements démographiques survenus au cours d'une année sont enregistrés dans différentes banques d'information. À des fins statistiques, les bébés québécois nés vivants sont inscrits dans le fichier des naissances du Bureau de la statistique du Québec. Les bébés ne présentant aucun signe de vie à la naissance et pesant au moins 500 grammes sont décrits dans le fichier des mortinaissances. Ces fichiers contiennent les données nécessaires au calcul et au suivi de nombreux indicateurs démographiques et de santé. Pour assurer la qualité de ces indicateurs, les renseignements compilés doivent être complets et exacts.

Or les fichiers québécois comportaient traditionnellement une proportion élevée de valeurs manquantes pour le poids du nouveau-né et le nombre de semaines de gestation, comparativement à ceux des autres provinces et à ceux d'autres pays. Au Québec en 1992, la durée de gestation était absente du formulaire de naissance dans 3,2 % des cas, et le poids à la naissance dans 1,1 % des cas. Pour les mortinaissances, ces proportions s'élevaient respectivement à 7,7 % et 10,4 %. Dans plusieurs pays, les valeurs manquantes représentent moins de 1 % des données pour ces variables (Notzon et Evans, 1992). Dans toutes les autres provinces canadiennes, elles représentaient en

* Respectivement du ministère de la Santé et des Services sociaux, Service du développement d'indicateurs, du ministère de la Santé et des Services sociaux, Service de l'infocentre, et du Bureau de la statistique du Québec. Remerciements à Russell Wilkins (Statistique Canada), Anne Martel, Alain Saucier et François Camirand (MSSS), Daniel Parent, Jean Lachapelle et Louis Duchesne (Bureau de la statistique du Québec).

1992 moins de 0,3 % des dossiers de naissances pour le poids et pour la durée de gestation (Wilkins, communication personnelle).

L'utilisation de différentes méthodes de calcul de la prématurité et de l'insuffisance de poids à la naissance, deux indicateurs très utilisés dans le domaine de la périnatalité, est un exemple typique des inconvénients occasionnés par l'ampleur des données manquantes. Plusieurs personnes ne tiennent pas compte des données manquantes, comme cela se fait habituellement lorsque les fichiers sont plus complets. D'autres les répartissent de différentes façons (proportionnelle ou autre). Ce manque d'uniformité dans les méthodes engendre évidemment aussi des disparités dans les résultats.

D'autres indicateurs, de plus en plus utilisés dans les pays à faible mortalité infantile, sont affectés par l'ampleur des données manquantes, par exemple la mortalité fœto-infantile (périnatale, néonatale, etc.) selon le poids à la naissance et selon la durée de gestation. La qualité des données concernant le poids et la durée de gestation dans les fichiers des naissances et des mortinaissances devient alors capitale, puisque ces fichiers sont fusionnés avec le fichier des décès pour le calcul de ces indicateurs. Les statistiques de mortalité selon le poids et la durée de gestation sont une source très riche de renseignements quant à la qualité des soins et permettent la comparaison entre établissements, groupes d'établissements, régions et pays. D'un pays à l'autre, les bébés très prématurés ou de très faible poids ne sont pas déclarés uniformément, surtout s'ils décèdent dans les minutes ou les heures suivant leur naissance. La situation est encore plus hétérogène pour les mortinaissances. Ces disparités affectent beaucoup moins les taux de mortalité selon le poids à la naissance que les taux traditionnels de mortalité.

C'est dans ce contexte qu'a été prise la décision de réduire la proportion des valeurs manquantes pour le poids ou la durée de gestation dans les fichiers québécois des naissances et des mortinaissances de 1993. La solution devait, pour être efficace, permettre de corriger à la source le plus de dossiers possible, cela avant la fermeture annuelle des fichiers, afin d'offrir un produit plus complet à tous les utilisateurs. Elle devait être simple à produire, économique et respecter la confidentialité des renseignements personnels.

Plusieurs avenues ont été envisagées : inciter fortement les centres hospitaliers à remplir correctement le formulaire; ou

encore attribuer, à partir de la donnée présente, la valeur de la donnée manquante (exemple : au poids «x» correspond habituellement une durée de gestation de «y» semaines, pour un nouveau-né de sexe «z»). Une solution plus intéressante, suggérée par Russell Wilkins de Statistique Canada, a été retenue : le jumelage de fichiers. La grande majorité des nouveau-nés inscrits dans les fichiers des naissances ou des mortinaissances le sont également dans le fichier des hospitalisations (MED-ÉCHO) du ministère de la Santé et des Services sociaux (MSSS). Or le fichier MED-ÉCHO contient les variables de poids à la naissance et de durée de gestation. Il a donc été décidé de jumeler les fichiers des naissances et des mortinaissances du Bureau de la statistique du Québec (BSQ) et celui du MSSS. L'objectif est d'identifier, dans le fichier des hospitalisations, le dossier qui correspond au dossier incomplet du fichier des naissances ou des mortinaissances. La donnée manquante est alors remplacée par celle présente au dossier du fichier d'hospitalisation.

Le jumelage des fichiers des événements démographiques n'est pas sans précédent. La plupart des pays de l'OCDE (Organisation de coopération et de développement économiques) procèdent au jumelage de banques de données pour créer des indicateurs dans le domaine de la santé. Par exemple, le jumelage du fichier des naissances avec le fichier des décès permet de calculer la mortalité spécifique selon le poids. Les cas documentés de jumelage de fichiers pour améliorer la qualité de la déclaration de l'un d'entre eux sont cependant rares. Le but de cette note de recherche est de décrire le jumelage du fichier des naissances et du fichier des hospitalisations, ainsi que son impact sur la proportion de données manquantes. Suivra une brève description de l'effet de ce processus lorsqu'il est appliqué au fichier des mortinaissances.

MÉTHODOLOGIE

La qualité du jumelage des dossiers repose sur une prémisse incontournable : il faut retrouver le dossier du même nouveau-né dans chacun des deux fichiers. On considère qu'il s'agit du même individu si, dans les deux sources, l'enfant a la même date de naissance, est né dans le même établissement et réside au même code postal, et sa mère a la même date de naissance.

Création d'un sous-fichier des naissances

Avant la fermeture du fichier annuel des naissances, un sous-fichier est créé pour le jumelage. Il comprend neuf variables : le *type de naissance* (simple ou multiple); les quatre variables servant de clé de fusion, soit le *code de l'établissement* où a lieu la naissance, la *date de naissance de la mère*, le *code postal de résidence* à six positions et la *date de naissance du nouveau-né*; les deux variables pour lesquelles on veut faire diminuer la proportion de données manquantes, soit la *durée de gestation* et le *poids à la naissance*; et les deux variables qui permettent de retracer dans le fichier maître des naissances les enregistrements à corriger à la fin du jumelage, soit le *numéro matricule du dossier* et le *numéro du formulaire*. Sont exclues les naissances des non-résidents du Québec et les naissances survenues hors des établissements québécois, car leur contrepartie n'est pas dans MED-ÉCHO.

Création d'un sous-fichier de MED-ÉCHO

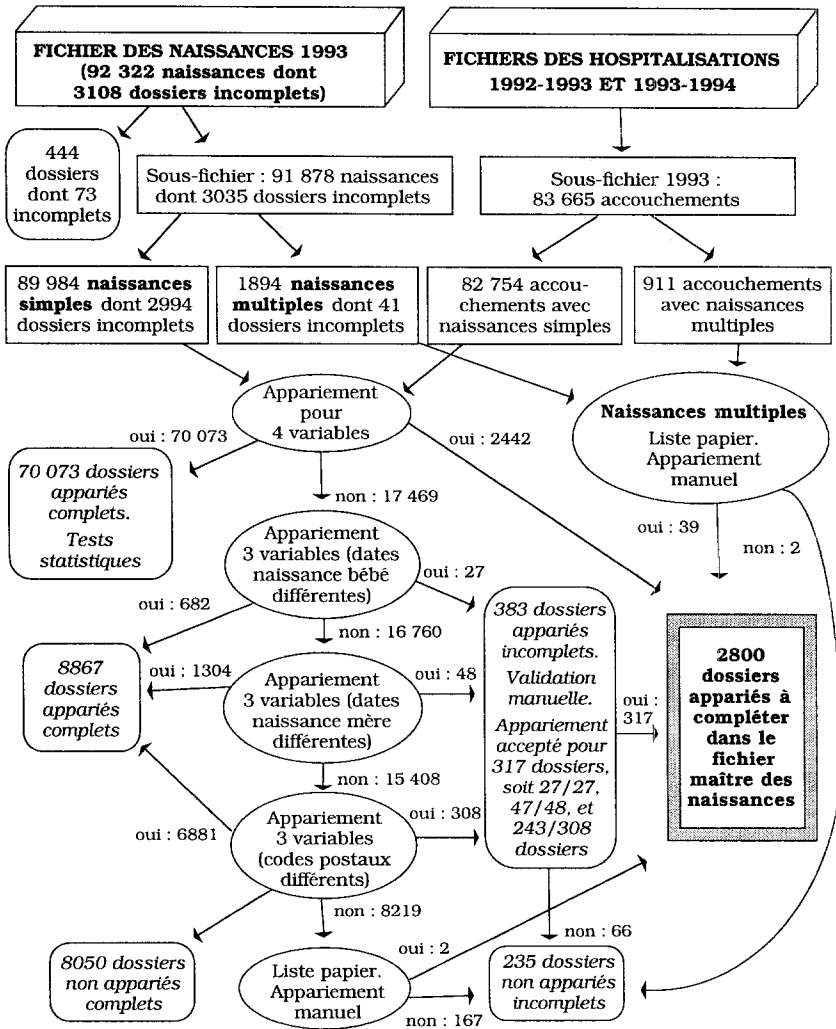
La création d'un sous-fichier de MED-ÉCHO est possible à partir du dossier de la mère ou du dossier du nouveau-né. Un pré-test a démontré les avantages d'utiliser le dossier de la mère. Les accouchements survenus de janvier à mars proviennent d'un fichier fermé et validé. Les dossiers d'avril à décembre proviennent d'un fichier en formation, donc incomplet et partiellement valide. Les quatre variables servant de clé de fusion, les deux variables à corriger et le type d'accouchement sont retenus.

Dans MED-ÉCHO, la date de naissance du nouveau-né n'est pas spécifiée dans le dossier de la mère. Elle est donc estimée à partir de la date du premier acte médical d'obstétrique noté au dossier au cours de l'hospitalisation de la mère, avec un écart accepté de plus ou moins deux jours. Il est en effet très fréquent que le premier acte (rupture de membranes, électrodes fœtales, péridurale, etc.) soit posé la veille de la naissance. L'écart peut être dans l'autre sens si le ou les seuls actes déclarés au dossier (réparation de déchirure etc.) sont postérieurs à la naissance.

Jumelage des dossiers des naissances multiples

L'étape suivante consiste à isoler dans chacun des sous-fichiers les naissances multiples des naissances simples

(figure 1). Les dossiers avec naissances multiples des deux sources sont intégrés en une seule liste. L'appariement vise à retrouver le dossier de MED-ÉCHO correspondant à chacun des dossiers qui, dans le fichier des naissances, comportent une donnée manquante pour le poids ou la durée de gestation.



Note : Les nombres correspondent aux dossiers du fichier des naissances, sauf s'il est spécifié que ce sont des accouchements.

FIGURE 1 — Étapes du jumelage des fichiers, 1993

Le jumelage doit être fait manuellement puisque deux naissances ou plus correspondent à un seul accouchement. Les poids d'au plus trois nouveau-nés sont inscrits dans le dossier de la mère. Dans les cas peu fréquents où, parmi les jumeaux ou triplés d'une famille, plus d'un poids à la naissance est absent, un jugement est porté pour chacun des dossiers. Par exemple, le poids supérieur est attribué au garçon plutôt qu'à la fille, à l'enfant né vivant plutôt qu'au mort-né, etc.

Jumelage des dossiers de naissances simples

Appariement des dossiers avec les quatre variables de la clé de fusion

Les naissances multiples étant traitées séparément, il ne reste que les dossiers des naissances simples. Une première fusion des deux sous-fichiers permet d'identifier tous les nouveau-nés qui, dans les deux fichiers, ont la même date de naissance, la même date de naissance de la mère, le même code postal de résidence et le même établissement de naissance. Les dossiers sont alors divisés en trois groupes.

Le premier groupe est composé des dossiers appariés dont le poids et la durée de gestation sont présents dans le fichier des naissances. Il fait l'objet de quelques tests statistiques pour vérifier la concordance des données pour le poids et la durée de gestation dans les deux sources. Une liste papier de ces dossiers appariés permettra en outre de valider l'exactitude de certaines données aberrantes dans le fichier des naissances, et de les corriger au besoin.

Le second groupe comprend les dossiers appariés, mais dont le poids ou la durée de gestation n'est pas inscrit dans le fichier des naissances.

Tous les dossiers déjà appariés sont mis de côté. Les autres tentatives de jumelage se concentrent sur le troisième groupe de dossiers, soit les naissances simples qui n'ont pas encore trouvé leur contrepartie dans MED-ÉCHO.

Appariement avec une clé partielle (trois variables sur quatre)

La difficulté de se souvenir de son code postal, un déménagement récent, la fièvre liée à l'événement, la barrière des langues ainsi que les erreurs d'inscription ou de saisie des données peuvent expliquer une partie des méprises dans l'un

ou l'autre des fichiers. Pour contourner ce problème, les trois jumelages suivants visent à identifier les dossiers qui correspondent au même nouveau-né dans les deux sources, mais pour lesquels une erreur relative à l'une des variables de la clé de fusion s'est glissée dans l'un ou l'autre des fichiers. Les valeurs des trois autres variables de la clé de fusion doivent être identiques dans les deux fichiers. Trois tentatives successives de jumelage partiel permettent d'apparier plusieurs dossiers du fichier des naissances avec des dossiers de MED-ÉCHO.

Trois listes sur support papier sont produites. Une vérification manuelle de chacun des dossiers permet de juger de l'exactitude du raccordement : les dossiers se rapportent-ils ou non au même individu ? Si oui, la donnée manquante est remplacée par la donnée de MED-ÉCHO. La donnée présente dans les deux fichiers (le poids à la naissance ou la durée de gestation) doit en outre concorder pour qu'on accepte de conclure que les deux enregistrements se rapportent au même individu. Dans les cas douteux, la consultation du formulaire «Bulletin de naissance vivante»¹ facilite la décision.

Appariement manuel

Finalement, une dernière liste papier comporte les résidus des deux fichiers n'ayant pu être appariés, et est soumise à un examen manuel. Les dossiers appariés et acceptés sont complétés dans le fichier maître des naissances.

RÉSULTATS

L'exemple de l'année 1993

Le jumelage du fichier des naissances et du fichier MED-ÉCHO a été fait pour les années 1993, 1994 et 1995 (en cours au moment d'écrire ces lignes). La figure 1 donne un aperçu du nombre de dossiers impliqués à chaque étape, pour l'année 1993. Les fichiers de 1992-1993 et de 1993-1994 de MED-ÉCHO comprennent près de 2,5 millions d'hospitalisations dont sont extraits 83 665 dossiers avec accouchement pour l'année 1993.

¹ Le formulaire «Bulletin de naissance vivante» s'appelait, avant 1994, «Déclaration de naissance vivante».

Parmi les 92 322 dossiers du fichier des naissances du Québec, 444 (0,5 %) ne sont pas soumis au jumelage (naissances hors du Québec etc.). La majorité des dossiers, soit 72 515 ou 78,5 %, sont appariés avec leur contrepartie dans MED-ÉCHO lors du premier jumelage avec les quatre clés de fusion. Les trois tentatives de jumelage partiel, avec trois clés sur quatre, permettent d'apparier 9250 autres dossiers (10,0 %), laissant 8219 dossiers (8,9 %) de naissances simples du fichier des naissances non appariés informatiquement. S'y ajoutent les 1894 dossiers de naissances multiples (2,1 %) qui, lorsqu'ils sont incomplets, sont soumis à un jumelage manuel.

Au départ, 3108 dossiers sont incomplets dans le fichier des naissances. Près de 90,1 %, soit 2800 dossiers, sont complétés avec le poids ou la durée de gestation trouvé dans le fichier MED-ÉCHO : 2442 dossiers (78,6 %) à partir de la fusion à quatre variables, 317 dossiers (10,2 %) à partir de celle à trois variables, deux dossiers après appariement manuel et 39 (1,3 %) provenant de la liste des jumeaux. Les 308 enregistrements non complétés sont composés des 73 dossiers (2,3 %) non soumis au jumelage et des 235 (7,6 %) qui résistent à l'appariement.

L'examen des dossiers incomplets, jumelés avec trois variables, montre que les erreurs dans le code postal sont de loin les plus nombreuses, avec 308 dossiers. Quelque 48 dossiers sont jumelés malgré une différence dans la date de naissance de la mère, et 27 autres concordent sauf pour la date de naissance du nouveau-né plus ou moins deux jours. La majorité (317) des appariements pour trois des quatre variables est acceptée après l'examen manuel des listes. Dans la plupart des dossiers, l'erreur est minime et le verdict facile à rendre : un 3 au lieu d'un 8, une inversion de chiffres (jour, mois), etc. Ce sont essentiellement les dossiers dont le code postal est très différent d'un fichier à l'autre qui ont été rejetés.

Parmi les 308 dossiers incomplets à la fin du processus, 95 enregistrements sont incomplets pour le poids à la naissance, 172 pour la durée de gestation et 41 pour les deux variables à la fois.

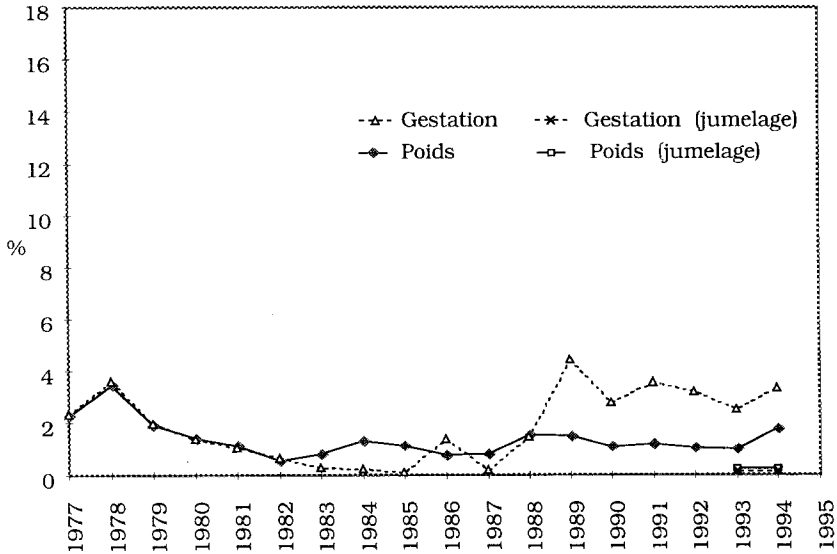
Un fichier des naissances plus complet

Le pourcentage de données manquantes quant au poids à la naissance et à la durée de gestation dans le fichier des naissances chute radicalement avec ce travail. La figure 2 montre

TABLEAU 1 — Fréquence des données manquantes pour le poids à la naissance et la durée de gestation, fichier des naissances et fichier des mortinaissances, Québec, 1991-1995

	Naissances vivantes			Mortinaissances		
	Poids	Durée de gestation	Total des naissances	Poids	Durée de gestation	Total des mortinaissances
	%	%	N	%	%	N
1991	1,2	3,6	97 348	11,0	3,5	456
1992	1,1	3,2	96 054	10,4	7,7	413
1993	0,1	0,2	92 322	0,3	0,3	380
1994	0,1	0,2	90 417	3,3	1,4	359
1995	n.d.	n.d.	n.d.	2,5	1,4	366

Source : Bureau de la statistique du Québec, fichiers des naissances et des mortinaissances.



Source : Bureau de la statistique du Québec, fichier des naissances.

Note : Pour 1993, les pourcentages avant jumelage sont des estimations prudentes.

FIGURE 2 — Pourcentage des données manquantes pour le poids et la durée de gestation dans le fichier des naissances, avant et après le jumelage, Québec, 1977-1994

que jamais entre 1977 et 1992 le pourcentage de données manquantes pour le poids à la naissance n'a été inférieur à 0,5 %. Il a varié entre 1,1 % et 1,5 % au cours de la période 1988-1992. Le jumelage avec le fichier des hospitalisations permet de ramener le taux à 0,1 % tant en 1993 qu'en 1994, alors qu'il aurait été, sans jumelage, de 1,8 % en 1994.

L'amélioration des données sur la durée de gestation est encore plus saisissante. Le pourcentage de données manquantes a atteint un sommet de 4,4 % en 1989 et est retombé à 0,2 % en 1993 et en 1994 après le jumelage. Sans jumelage, il aurait été de 3,3 % en 1994.

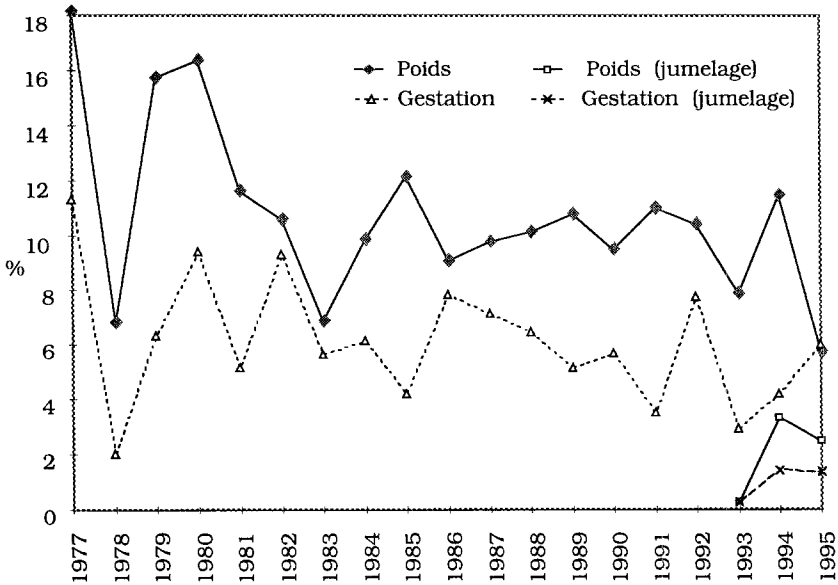
Entre 1991 et 1993, le nombre de données manquantes pour le poids à la naissance est donc passé de 1161 à 136 et celui des données manquantes pour la durée de gestation, de 3463 à 213.

Un fichier des mortinaissances plus complet

Le même processus est appliqué au fichier des mortinaissances, avec les mêmes variables et avec un succès encore plus spectaculaire que dans le cas des naissances. Le jumelage se fait manuellement, à partir de listes sur support papier, en raison du petit nombre de cas. Des 41 dossiers incomplets au chapitre de la durée de gestation ou du poids à la naissance en 1993, 40 sont appariés avec leur contrepartie dans MED-ÉCHO et complétés. La figure 3 montre que le problème était beaucoup plus aigu dans le fichier des mortinaissances que dans le fichier des naissances, et ce depuis longtemps. Les pourcentages de données manquantes pour les mortinaissances étaient de 10,4 % pour le poids à la naissance et de 7,7 % pour la durée de gestation en 1992. Le jumelage du fichier des mortinaissances de 1993, effectué en mai, se conclut par 0,3 % de données manquantes pour le poids et 0,3 % pour la durée de gestation. L'année suivante, le jumelage pour l'année 1994 est fait un mois plus tôt et laisse davantage de données manquantes, soit respectivement 3,3 % et 1,4 % (tableau 1). Les fichiers des naissances et des mortinaissances se comparent désormais avantageusement aux fichiers des autres provinces.

Correction de valeurs aberrantes

Quelques dossiers du fichier des naissances ou du fichier des mortinaissances sont corrigés pour le poids ou la durée de



Source : Bureau de la statistique du Québec, fichier des mortinaissances.

FIGURE 3 — Pourcentage des données manquantes pour le poids et la durée de gestation dans le fichier des mortinaissances, avant et après le jumelage, Québec, 1977-1995

TABLEAU 2 — Proportion des dossiers dans lesquels le poids à la naissance ou la durée de gestation est absent dans le fichier canadien des naissances, selon les provinces, 1992 et 1994 (%)

Province	Poids à la naissance		Durée de gestation	
	1992	1994	1992	1994
Terre-Neuve	0,03	0,11	0,09	0,02
Île-du-Prince-Édouard	0,22	0,24	0,00	0,00
Nouvelle-Écosse	0,15	0,01	0,17	0,00
Nouveau-Brunswick	0,00	0,00	0,02	0,01
Québec	1,02	0,14	3,19	0,23
Ontario	0,27	0,26	0,17	0,29
Manitoba	0,01	0,02	0,02	0,02
Saskatchewan	0,00	0,00	0,00	0,00
Alberta	0,00	0,00	0,00	0,00
Colombie-Britannique	0,04	0,46	0,04	0,43
Yukon	0,00	0,00	0,00	0,00
Territoires du Nord-Ouest	0,08	0,60	0,08	0,00
Canada	0,36	0,20	0,83	0,22

Source : Russell Wilkins, Statistique Canada.

gestation. C'est le cas des dossiers de nouveau-nés dont le poids est improbable en raison de la durée de gestation. Après vérification dans le dossier correspondant de MED-ÉCHO, la donnée erronée est alors identifiée et corrigée. Parmi les dossiers fusionnés avec trois des quatre variables de la clé de fusion, l'opération permet aussi de compléter ou de corriger au passage près d'une centaine de dossiers ayant des données manquantes ou erronées quant au code postal, à la date de naissance de la mère ou à celle du nouveau-né.

Détection d'erreurs systématiques dans MED-ÉCHO

Le processus permet également de déceler des erreurs dans MED-ÉCHO. Une erreur systématique dans la déclaration du poids à la naissance par un centre hospitalier dans le fichier non fermé de MED-ÉCHO de 1994-1995 a ainsi été découverte et corrigée.

Appréciation de la qualité des données

L'exercice permet d'apprécier la qualité de la déclaration du poids à la naissance et de la durée de gestation dans les fichiers des naissances et MED-ÉCHO. Au début du processus de jumelage, avant que l'on effectue quelque correction que ce soit, la grande majorité des dossiers dans les deux fichiers ont été appariés pour les quatre variables de la clé (70 073 naissances) et étaient complets quant au poids et à la durée de gestation dans les deux fichiers. On a analysé ces dossiers pour vérifier la concordance des valeurs inscrites dans chacune des sources. Ainsi, 83,9 % ont exactement la même durée de gestation inscrite dans les deux fichiers, 13,6 % ont une semaine de différence, 1,8 % ont deux semaines de différence et 0,7 % ont un écart de trois semaines et plus.

Pour le poids à la naissance, 95,1 % des nouveau-nés ont le même poids à 50 grammes près dans les deux fichiers, 3,4 % présentent une différence de 51 à 150 grammes, et 1,5 % une différence de poids supérieure à 150 grammes.

Il est très difficile de savoir dans quel fichier la donnée est erronée, sauf si l'écart entre les valeurs est très important (exemple : un dossier décrivant un nouveau-né pesant 500 grammes après une gestation de 40 semaines). L'accès au formulaire «Bulletin de naissance» permet à l'occasion d'identifier une erreur de saisie.

RESPECT DE LA CONFIDENTIALITÉ

Les travaux sont réalisés en concertation avec trois partenaires : la copilote des orientations du fichier des événements démographiques au Service du développement d'indicateurs du MSSS, la responsable de la qualité des fichiers des événements démographiques au Bureau de la statistique du Québec et le pilote opérationnel de ces fichiers au service de l'Infocentre du MSSS. Deux programmeurs contribuent également au projet, l'une à l'Infocentre, l'autre au BSQ. Cinq personnes ont donc accès aux listes de données issues du jumelage.

Seules les personnes du Bureau de la statistique du Québec ont accès aux formulaires du fichier des événements démographiques, donc aux noms et adresses complets des personnes. Les autres partenaires du projet n'y ont pas accès. Personne n'a accès aux noms et adresses des personnes inscrites au système MED-ÉCHO puisque les centres hospitaliers ne les transmettent pas au MSSS.

Par contre, les données sont assez précises (dates de naissance de la mère et du nouveau-né, code postal à six positions, code de l'établissement de naissance du nouveau-né) pour permettre d'identifier les individus. Des mesures strictes sont prises pour que fichiers et listes papier ne soient accessibles qu'aux cinq personnes impliquées dans le projet de jumelage. Dans le cadre de leurs tâches régulières de vérification ou de saisie de données, quelques personnes du BSQ doivent traiter les résultats d'appariements, sans avoir accès aux dossiers de MED-ÉCHO.

LIMITES

Le but de cette opération est de remplacer, dans le plus grand nombre possible de dossiers, la donnée manquante dans le fichier des naissances par la donnée se rapportant au même individu dans le fichier MED-ÉCHO. Le risque d'introduction d'une donnée erronée dans le fichier des naissances et l'incapacité de compléter certains dossiers sont les deux principales limites de cette démarche.

Il peut y avoir confusion dans l'identification du nouveau-né, d'où l'attribution d'un poids ou d'une durée de gestation erroné. Le risque est plus élevé dans les cas où une des variables de la clé de fusion ne concorde pas. Chaque année, deux personnes vérifient avec attention les dossiers litigieux. Si le

poids ou la durée de gestation présent dans les deux fichiers concorde, et si l'erreur semble banale (inversion de chiffres), l'appariement est automatiquement accepté. C'est le cas de la quasi-totalité des dossiers qui ne concordent pas pour la date de naissance de la mère ou du nouveau-né.

La majorité des dossiers pour lesquels le code postal est différent n'ont qu'une légère discordance. Plusieurs diffèrent cependant pour trois positions ou plus sur six. Ces dossiers posent davantage problème. La première année, beaucoup d'énergie leur a été consacrée. Deux questions étaient posées. L'examen de la liste des dossiers non appariés permet-il d'identifier tout autre dossier dans MED-ÉCHO avec lequel il aurait pu y avoir confusion, autrement dit, cette année-là, un autre enfant dont la mère aurait la même date de naissance serait-il né ce jour-là dans le même hôpital ? La consultation du formulaire «Bulletin de naissance» peut-elle permettre d'identifier une erreur de saisie, ou donner l'indice d'un déménagement entre la date de naissance et le moment où le formulaire est envoyé (ratures, etc.) ?

Une attitude prudente a conduit, la première année, au rejet de l'appariement d'un dossier sur cinq parmi les dossiers présentant une différence dans le code postal. Elle a cédé le pas, les années subséquentes, à une attitude plus libérale amenant l'acceptation de la quasi-totalité des dossiers, dans la mesure où il n'y avait pas de disparité entre la donnée sur le poids ou la durée de gestation déjà présente dans le fichier des naissances et celle du fichier MED-ÉCHO. Le code postal semble être une information difficile à retenir et à écrire correctement !

Le fichier MED-ÉCHO n'est pas exempt d'erreurs de déclaration ou de saisie. Bien que, pour la majorité des nouveau-nés, le poids et la durée de gestation aient la même valeur dans le fichier des naissances et dans le fichier MED-ÉCHO, il reste que la durée de gestation diffère de deux semaines ou plus et le poids à la naissance de 150 grammes ou plus pour respectivement 2,5 % et 1,5 % des dossiers. L'erreur peut provenir de l'un ou l'autre des fichiers. Il est donc possible qu'une donnée manquante soit remplacée par une donnée erronée. Le poids n'est pas soumis à de larges validations dans MED-ÉCHO, sauf dans le cas des nouveau-nés de très faible poids : la destination de ces enfants ne peut être le domicile après quelques jours d'hospitalisation. Avant que la donnée de MED-ÉCHO soit incluse dans le fichier des naissances, une validation permet

d'écarter les dossiers qui amèneraient l'introduction d'une donnée aberrante (poids eu égard à la durée de gestation).

Le poids à la naissance est exprimé de façon moins précise dans MED-ÉCHO. Alors que, dans le fichier des naissances, le poids est inscrit au gramme près, il est exprimé dans MED-ÉCHO par tranches de 100 grammes, en raison du manque d'espace dans le fichier. Par exemple, un poids de 2365 grammes est inscrit tel quel dans le fichier des naissances. Dans MED-ÉCHO, la valeur est habituellement tronquée (exemple : 23), parfois arrondie (exemple : 24). Cinquante grammes sont donc ajoutés aux poids de MED-ÉCHO lorsqu'ils remplacent une donnée manquante dans le fichier des naissances (exemple : 2350). Les valeurs transcrites dans le fichier des naissances sont donc un peu moins précises que les autres.

Le jumelage des fichiers ne pourra jamais permettre de corriger tous les dossiers. Plusieurs groupes ne se retrouvent pas dans MED-ÉCHO, tels les enfants nés à domicile ou dans une maison des naissances au Québec, et les Québécois et Québécoises nés hors du Québec (cas fréquent dans les régions frontalières comme l'Outaouais). Certains dossiers correspondant au même nouveau-né mais pour lesquels il y a une erreur dans plus d'une des variables de la clé de fusion ne peuvent pas être complétés en raison du doute raisonnable qui persiste.

La date du jumelage influence le pourcentage des dossiers complétés : plus on attend pour le faire, plus le fichier MED-ÉCHO de la dernière année financière est complet, plus le nombre d'appariements est élevé et plus les corrections sont nombreuses. Pour minimiser le nombre de données manquantes, il faudrait idéalement effectuer le jumelage tout juste avant la fermeture de MED-ÉCHO. Cela heurte l'impératif de fermer dans un délai raisonnable les fichiers des événements démographiques. Les utilisateurs veulent évidemment qu'ils soient disponibles le plus rapidement possible. Un arbitrage doit donc être fait chaque année pour déterminer la date du jumelage. Pour gagner en efficacité, le jumelage a été placé à la toute fin du processus de validation et de fermeture des fichiers des naissances et des mortinaissances au BSQ.

CONCLUSION

L'opération de jumelage des fichiers des naissances et des mortinaissances avec le fichier des hospitalisations (MED-

ÉCHO) a permis de faire chuter de façon importante le nombre de données manquantes dans les fichiers des naissances et des mortinaissances. Le processus permet en outre de croire que les données complétées sont exactes, puisqu'elles font référence au même nouveau-né, dans un fichier différent. Plus importantes la première année lors du développement des programmes informatiques et du rodage du processus de validation des données, les ressources nécessaires au jumelage sont dorénavant modestes. Cette opération fait désormais partie intégrante du processus de fermeture annuelle des fichiers des naissances et des mortinaissances.

Le jumelage des fichiers améliore la qualité des indicateurs. Il met un terme aux diverses méthodes de calcul du taux d'insuffisance de poids à la naissance et du taux de prématurité. Il est désormais inutile de répartir les données manquantes : il suffit de ne pas en tenir compte. Le calcul des taux de mortalité selon le poids à la naissance ou la durée de gestation sera plus facile à réaliser et plus fiable que lorsque le nombre de données manquantes était élevé.

L'expérience est donc concluante en ce qui a trait tant aux naissances qu'aux mortinaissances et elle démontre que le jumelage de fichiers est une façon économique et efficace d'assurer une meilleure qualité des données et des indicateurs de santé qui en découlent.

Des travaux en cours prolongent cet effort d'amélioration de la qualité des fichiers des naissances, des mortinaissances et des hospitalisations, en développant une méthode permettant de corriger les dossiers erronés. Ils s'articulent autour des dossiers dont les valeurs de poids et de durée de gestation s'écartent significativement d'un fichier à l'autre, et des dossiers pour lesquels la valeur du poids à la naissance semble aberrante compte tenu de la durée de gestation.

RÉFÉRENCES

- NOTZON, Francis, et Stephen EVANS, 1992. «Differences and trends in birthweight distributions among the ICE countries», *Proceedings of the International Collaborative Effort on Perinatal and Infant Mortality*. Maryland, NCHS, vol. III : V3-V27.
- WILKINS, Russell (Statistique Canada), communication personnelle.

RÉSUMÉ — ABSTRACT — RESUMEN

MONTREUIL Sylvie, HALLEY Richard et JOE Shirley — POIDS À LA NAISSANCE ET DURÉE DE GESTATION MANQUANTS ? LA SOLUTION DANS LE JUMELAGE DES FICHIERS DES NAISSANCES ET DES HOSPITALISATIONS

Au Québec, le nombre élevé de données manquantes pour le poids à la naissance et la durée de gestation dans le fichier des naissances engendrait des problèmes dans le calcul de certains indicateurs de santé. Or ces données sont présentes dans le fichier des hospitalisations. Le jumelage des deux fichiers permet de trouver le dossier se rapportant au même individu dans chaque fichier, de compléter les données manquantes et d'améliorer la qualité du fichier des naissances. Le jumelage s'est révélé un succès puisqu'il a permis de diminuer la proportion des valeurs manquantes dans le fichier des naissances à 0,1 % pour le poids à la naissance et à 0,2 % pour la durée de gestation en 1993, comparativement à 1,1 % et 3,2 % respectivement en 1992. Le même travail a été fait pour le fichier des mortinaissances, avec un résultat encore plus spectaculaire : les pourcentages de valeurs manquantes sont alors passés de 10,4 % et 7,7 % respectivement en 1992 à 0,3 % dans les deux cas en 1993.

MONTREUIL Sylvie, HALLEY Richard and JOE Shirley—MISSING DATA ON BIRTH WEIGHT AND GESTATIONAL AGE? A SOLUTION BY MATCHING BIRTH AND HOSPITALIZATION RECORDS

The high percentage of missing data on birth weight and gestational age in Quebec's birth registration file made it difficult to calculate certain health indicators. However, these data were available in the hospitalization files. Matching the two files made it possible to find the same individual's record in each file, fill in the missing data and improve the quality of the birth records. The matching proved successful in that it reduced the percentage of missing values in the birth registration file to 0.1% for birth weight and 0.2% for gestational age in 1993, compared to 1.1% and 3.2% respectively in 1992. The same procedure was used for stillbirth records, with even more dramatic results: the percentages of missing values fell from 10.4% and 7.7% respectively in 1992 to 0.3% in both categories in 1993.

MONTREUIL Sylvie, HALLEY Richard y JOE Shirley — PESO AL NACIMIENTO Y DURACIÓN DE GESTACIÓN FALTANTES : LA SOLUCIÓN EN EL EMPAREJAMIENTO DE LOS REGISTROS DE LOS NACIMIENTOS Y DE LAS HOSPITALIZACIONES

En Quebec, puesto que faltaba un número elevado de datos en cuanto al peso al nacimiento y la duración de gestación en el registro de nacimientos, resultaba difícil calcular ciertos indicadores de salud. Ahora bien, estos datos se encuentran presentes en el registro de

RÉSUMÉ — ABSTRACT — RESUMEN

hospitalizaciones. El emparejamiento que se propone de ambos registros permite encontrar el expediente de un mismo individuo en cada registro, completar los datos faltantes y mejorar la calidad del registro de los nacimientos. El emparejamiento ha resultado un éxito puesto que ha permitido reducir la proporción de valores faltantes en el registro de nacimientos a 0.1% para el peso al nacimiento, y a 0.2% para duración de gestación en 1993, comparativamente a 1.1% y 3.2% respectivamente en 1992. Se ha realizado el mismo trabajo para el registro de las mortinatalidades, con un resultado aún más espectacular: los porcentajes de valores faltantes pasaron en este caso de 10.4% y 7.7% respectivamente en 1992 a 0.3% en ambos casos en 1993.