

Les fichiers statistiques des naissances, mariages et décès du ministère des Affaires sociales (MAS) : quelques notes sur la qualité des fichiers et des données

Jacques Locas

Volume 11, numéro 2, août 1982

URI : <https://id.erudit.org/iderudit/600877ar>

DOI : <https://doi.org/10.7202/600877ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association des démographes du Québec

ISSN

0380-1721 (imprimé)

1705-1495 (numérique)

[Découvrir la revue](#)

Citer cette note

Locas, J. (1982). Les fichiers statistiques des naissances, mariages et décès du ministère des Affaires sociales (MAS) : quelques notes sur la qualité des fichiers et des données. *Cahiers québécois de démographie*, 11(2), 277–288.
<https://doi.org/10.7202/600877ar>

LES FICHIERS STATISTIQUES DES NAISSANCES, MARIAGES
ET DÉCÈS DU MINISTÈRE DES AFFAIRES SOCIALES (MAS):
QUELQUES NOTES SUR LA QUALITÉ DES FICHIERS ET DES DONNÉES

Jacques LOCAS*

Note présentée au 50^e congrès de l'ACFAS,
section de démographie, mai 1982

Cette communication porte sur trois éléments permettant d'évaluer la qualité des données des fichiers statistiques des naissances, mariages et décès du ministère des Affaires sociales: l'intégralité des fichiers statistiques, l'intégrité des données et des fichiers statistiques et l'actualité des données présentes aux différents fichiers¹. Avant d'aborder ces trois éléments, il importe de donner quelques définitions de base.

1. Définitions

- 1.1 Fichier-maître: fichier sur lequel sont enregistrées les informations telles qu'elles apparaissent sur les formulaires de déclaration du ministère des Affaires sociales; le fichier-maître contient également les informations obtenues par correspondance, quand il s'avère que les données contenues aux formulaires sont incomplètes, erronées ou incohérentes.
- 1.2 Fichier statistique: copie du fichier-maître à laquelle des corrections ont été apportées au cours de l'opération de clôture de l'année statistique (M. Rochon, 1979, K. Vovan, 1979).
- 1.3 Événements inscrits aux fichiers statistiques: seuls les naissances et les décès de résidents du Québec, que ces événements

* Service de la gestion et de l'exploitation des systèmes, ministère des Affaires sociales, 1279, boulevard Charest Ouest, Québec, QC, G1N 2C9.

1. Tous les résultats proviennent d'études ponctuelles que nous avons effectuées au Service des registres du ministère des Affaires sociales.

soient survenus au Québec ou hors Québec, sont présents aux fichiers statistiques et seuls les mariages dont le lieu de célébration est situé au Québec, qu'il s'agisse de résidents ou de non-résidents du Québec, sont inscrits au fichier statistique des mariages.

- 1.4 Année statistique: il s'agit de la période de référence durant laquelle les événements survenus durant une année civile "t" sont déclarés et dont les déclarations parviennent au MAS avant le 1er avril de l'année "t + 1". L'année statistique s'étend donc du 1er janvier de l'année civile "t" jusqu'au 31 mars de l'année suivante "t + 1".

Tous les événements présents aux différents fichiers statistiques, à partir de 1980, sont ceux dont la déclaration est parvenue au MAS pendant l'année statistique, période de référence des fichiers statistiques.

- 1.5 Déclaration "tardive" ou "différée": déclaration d'un événement survenu au cours de l'année "t", mais qui parvient au MAS après la fin de l'année statistique, c'est-à-dire après le 31 mars de l'année "t + 1". Les déclarations tardives ne seront présentes, après la fin de l'année statistique, qu'aux fichiers-maîtres qui, eux, sont cumulatifs et peuvent être mis à jour.

2. L'intégralité des fichiers

On peut définir l'intégralité d'un fichier comme étant l'état hypothétique d'un fichier où tous les événements survenus au cours d'une année civile "t" auraient été déclarés et dont toutes les déclarations seraient parvenues au MAS avant la fin de l'année statistique. Situation idéale, il va sans dire, puisque certains événements ne seront jamais déclarés et que des déclarations ne seront jamais reçues ou ne parviendront qu'après la fin de l'année statistique.

L'intégralité des fichiers sera estimée en posant l'hypothèse qu'un an après la fin de l'année "t" (c'est-à-dire en "t + 1"), tous les événements survenus en "t" auront été inscrits aux différents fichiers statistiques. Le taux d'intégralité d'un fichier statistique sera alors le rapport du nombre d'événements survenus en "t" et dont la déclaration est reçue pendant l'année statistique (c'est-à-dire du 1er janvier de l'année "t" au 31 mars de l'année "t + 1") à celui du nombre d'événements survenus en "t" et dont la déclaration parvient au MAS entre le 1er janvier de l'année "t" au 31 décembre de l'année "t + 1".

Ainsi, les données contenues dans les différents fichiers statistiques pour l'année statistique 1980 nous permettent de supposer qu'au 31 mars 1981, sur une base de cueillette de déclarations de deux ans, près de 99,5% des déclarations d'événements survenus en 1980 étaient présentes aux différents fichiers statistiques. Les taux d'intégralité ratta-

chés à chacun des fichiers apparaissent au tableau 1².

Tableau 1
Nombre d'événements survenus en 1980 et
présents aux fichiers statistiques à
différentes dates, Québec, 1980

Nombre	Fichier statistique		
	Naissances	Mariages	Décès
d'événements au 31 mars 1981: (1)	97 498	44 849	43 515
de déclarations tardives reçues entre le 1er avril 1981 et le 31 décembre 1981: (2)	316	376	295
d'événements au 31 décembre 1981: (3) = (1) + (2)	97 814	45 225	43 810
Taux d'intégralité du fichier: (1) / (3)	99,68%	99,17%	99,33%

3. L'intégrité des fichiers statistiques et des données reliées aux variables

3.1 Définitions

Un deuxième élément d'appréciation des fichiers statistiques concerne l'aspect qualitatif de leur contenu; nous parlons alors de l'intégrité des fichiers et des données reliées aux variables qui permettront la mise en tableaux.

Idéalement, un fichier statistique serait intègre s'il était le reflet exact des déclarations contenues sur les formulaires avec leurs

2. Pour une évaluation de la situation antérieure à 1975 relativement à l'intégrité des fichiers, on consultera les textes de E. Gabbay, 1975, R. Maheu, 1975 et L. Roy, 1977.

omissions, leurs aberrations et leurs incohérences.

On mesure l'intégrité par le nombre de modifications ou de corrections apportées aux données, au cours du processus de clôture de l'année statistique, pour les rendre cohérentes entre elles au fichier statistique. On pose alors les hypothèses suivantes:

- le fichier-maître est intègre avant le processus de clôture de l'année statistique;
- toutes les corrections ou modifications apportées au cours du processus de clôture de l'année statistique constituent des altérations par rapport à l'état du fichier-maître avant le processus de clôture de l'année statistique.

On obtient alors un taux d'altération qui est le rapport du nombre de corrections apportées aux fichiers lors du processus de clôture de l'année statistique au nombre de corrections possibles en supposant que toutes les données devraient être corrigées. Le taux d'intégrité serait le complément du taux d'altération.

3.2 La correction des données

Au cours du processus de clôture de l'année statistique, des listes de correction signalent tous les cas qui, répondant à des critères définis, devront être vérifiés et corrigés, le cas échéant. A partir de ces listes, deux types de corrections peuvent être faits:

- le premier type de corrections, dites "réelles", consiste à retrouver sur la formule de déclaration de l'événement les données exactes mal inscrites au fichier-maître. Ces corrections altèrent à la fois le fichier-maître et le fichier statistique;
- le deuxième type de corrections, dites "fictives", remplace des données aberrantes ou inexactes par des données probables parce que les données exactes ne se retrouvent pas sur les formules de déclaration. Ce type de correction répond essentiellement à la nécessité d'une plus grande cohérence entre les données.

Des corrections aléatoires sont aussi faites au moment où l'on constitue le fichier statistique proprement dit, c'est-à-dire après que le fichier-maître ait été épuré. Ces corrections remplacent des données annulées lors de corrections fictives ou complètent des informations manquantes au fichier-maître par des données attribuées de manière aléatoire selon des modèles de distribution qui sont introduits aux programmes de création du fichier statistique. Ces corrections ne sont apportées qu'au fichier statistique.

3.3 L'intégrité des fichiers statistiques

En ce qui concerne l'intégrité des fichiers statistiques, on peut poser le problème de la façon suivante: si le fichier statistique,

pour être intègre, doit refléter exactement le contenu du fichier-maître avant le processus de clôture de l'année statistique, quel est le biais introduit au fichier statistique en tant que copie du fichier-maître, si les corrections réelles, fictives et aléatoires faites soit au fichier-maître ou au fichier statistique devaient être considérées comme des altérations au contenu des fichiers?

Un relevé exhaustif de toutes les corrections apportées aux différents fichiers lors du processus de clôture de l'année statistique 1980 a été fait; 32 variables étaient impliquées par la correction de données pour l'ensemble des fichiers: 14 pour les fichiers des naissances, 9 pour ceux des décès et 9, également, pour les fichiers des mariages (tableau 2).

10 877 corrections furent apportées à l'ensemble des fichiers sur un nombre possible de 2 160 238 (avec 32 variables). Nous obtenons ainsi un taux d'altération pour l'ensemble des fichiers de 5% et, pour chacun des fichiers, les taux suivants:

<u>Fichiers</u>	<u>Taux d'altération (%)</u>
Naissances	4,7
Mariages	1,4
Décès	9,8

C'est donc le fichier statistique des décès qui serait le moins intègre avec un taux d'altération de près de 1%.

Pour chacun des fichiers statistiques, les corrections aléatoires constituent l'altération la plus importante avec des proportions de corrections de 81%, 54 et 59% respectivement pour les fichiers des naissances, mariages et décès. Pour tous les fichiers, les corrections fictives représentent environ 14% des corrections totales.

3.4 L'intégrité des données rattachées aux variables

On fait ici la même hypothèse que précédemment, c'est-à-dire que toutes les corrections apportées aux données des différentes variables sont autant d'altérations à l'intégrité des variables.

Pour le fichier des naissances (tableau 3), les variables dont les données ont été le plus souvent corrigées sont le rang de naissance avec un taux d'altération de 35% (3 378 corrections), la date de naissance de la mère (777) et l'état civil de la mère avec des taux de 8% environ et la municipalité de résidence de la mère (5,4% et 530 corrections).

Par ailleurs, l'état civil du décédé est la variable qui a nécessité le plus d'altérations au fichier des décès avec un taux de 47,4 données sur 1 000 (2 062 corrections). Plus de la moitié des corrections

Tableau 2
Nombre total de corrections selon le type effectuées aux
fichiers des naissances, mariages et décès de 1980

Fichier	Type de corrections			Total	Variables (N)	Événements (N)	Corrections possibles (N)	Taux d'altération ^a	
	Réelles	Fictives	Aléatoires					H ₀	H ₁
Naissances	370 (0,06)	855 (0,13)	5 257 (0,81)	6 482 (0,596)	14	97 498	1 364 972	0,00063	0,0047
Mariages	186 (0,33)	74 (0,13)	304 (0,54)	564 (0,052)	9	44 849	403 641	0,00017	0,0014
Décès	1 034 (0,27)	549 (0,14)	2 248 (0,59)	3 831 (0,352)	9	43 515	391 645	0,0014	0,0098
Ensemble	1 590 (0,146)	1 478 (0,136)	7 809 (0,718)	10 877	32		2 160 238	0,00068	0,005

Note: a) Taux d'altération: nombre de corrections effectuées/nombre de corrections possibles;
ce taux est soumis aux hypothèses H₀ et H₁

H₀ = hypothèse voulant que seules les corrections fictives altèrent le (s) fichier (s)

H₁ = hypothèse voulant que toutes les corrections altèrent le (s) fichier (s)

Tableau 3
Nombre total de corrections et taux d'altération par variable
et par fichier pour l'année statistique 1980

Variable ^a	Naissances		Mariages			Décès		
	Corrections (N)	Taux d'altération ^b (‰)	Variable	Corrections (N)	Taux d'altération (‰)	Variable	Corrections (N)	Taux d'altération (‰)
R.N.	3 378	35,0 ^c	DNH	121	2,7	E.C.D.	2 062	47,4
ECM	783	8,0	DNF	108	2,4	I.A.	670	15,0
DNM	777	8,0				DND	340	7,8
MR	530	5,4				CAUSES	246	5,7
						MR	232	5,3
(...)	(...)		(...)	(...)		(...)	(...)	
TOTAL	6 482	4,47 ^d	TOTAL	564	0,9	TOTAL	3 831	7,0

Notes: a) Signification des sigles des variables: R.N.: rang de naissance, ECM: état civil de la mère, DNM: date de naissance de la mère, MR: municipalité de résidence, DNH: date de naissance de l'homme, DNF: date de naissance de la femme, ECD: état civil du décédé, I.A.: indicateur d'âge, DND: date de naissance du décédé.

b) Taux d'altération: nombre de corrections effectuées/nombre de corrections possibles.

c) Les dénominateurs des taux d'altération par variable sont selon le fichier des naissances: 97 498, des mariages: 44 849 et des décès: 43 515.

d) Les dénominateurs des taux d'altération pour l'ensemble des variables de chaque fichier sont pour le fichier des naissances: 1 364 972, le fichier des mariages: 403 641 et le fichier des décès: 391 645.

totales se rapporte à cette variable sur ce fichier. Les indicateurs d'âge et la date de naissance du décédé sont les deux autres variables les plus touchées avec des taux d'altération de 15% et 7,8%. La cause de décès et la municipalité de résidence du décédé ont chacune des taux de 5,5% environ.

Le fichier des mariages, par ailleurs, a surtout été altéré par la correction des dates de naissance de l'époux et de l'épouse, corrections qui représentent plus de 40% des corrections totales apportées au fichier et des taux d'altération de 2,5% environ.

Globalement, pour l'ensemble des fichiers (tableau 4), on observe donc que les corrections apportées aux données des variables relatives aux dates, à l'état civil et au rang de naissance représentent 77% de toutes les corrections apportées aux 32 variables.

4. L'actualité des données³

De 1976 à 1980, le délai moyen de constitution⁴ pour l'ensemble des fichiers est passé de 32 mois révolus à 5 mois seulement (tableau 5). Ce délai de constitution des fichiers statistiques se compose en deux étapes successives: une première étape consiste à améliorer l'intégralité des fichiers, opération qui s'effectue du 1er janvier de l'année "t + 1" au 31 mars de l'année "t + 1" et que l'on pourrait appeler "délai de récupération" n'ayant pas été expédiées durant l'année civile; une deuxième étape, ou "délai de production" proprement dit, met en cause la correction de données après la fin de l'année statistique; elle exige le soutien continu de ressources informatiques adéquates pour le traitement des données. C'est au cours de cette dernière étape que l'on traite de l'intégrité des fichiers et des données.

L'expérience des années récentes montre que si l'on désire diffuser les données le plus rapidement possible, c'est en diminuant le "délai de production", grâce à un soutien informatique continu, qu'on y parviendra. Il n'y a actuellement aucun gain à espérer de l'entreprise de récupération, les réformes quant à la cueillette des déclarations ayant déjà été réalisées suite à l'adoption de la Loi de la protection de la santé publique (1972).

On peut donc aisément comprendre ici que les deux premiers objectifs (l'intégralité et l'intégrité) entrent en conflit avec le troisième (l'actualité des données). En effet, plus on met de temps à s'assurer d'une

3. R.J. LOUP et B.V. THOMPSON (1982), dans un document interne de l'OMS, parlent aussi d'"opportunité" des données qu'ils définissent comme la nécessité que "toutes les données appropriées doivent être disponibles au moment voulu".

4. Temps écoulé entre la fin de l'année civile et la date de constitution du fichier statistique relatif à celle-ci.

Tableau 4
Nombre de corrections et taux d'altération par groupe de variable
pour l'année statistique 1980

Variable	Fichier			Total	Corrections possibles par groupe de variable	Taux d'alté- ration sur les groupes de variable (%)	Taux d'altération pour l'ensemble des variables (32) de tous les fichiers (%)
	Naissances	Mariages	Décès				
Dates	1 213 (5) ^a	410 (5)	372 (2)	1 995 (12)	798 765	2,5	0,09
Etat civil	783 (1)	111 (2)	2 062 (1)	2 956 (4)	230 711	12,8	1,4
Rang de nai- sance	3 378 (1)			3 378 (1)	97 498	34,6	1,6
TOTAL	5 374 (7)	521 (7)	2 434 (3)	8 329 (17)	1 126 974		
Taux d'alté- ration sur l'ensemble des variables d'un fichier (%)	3,9	1,3	6,2				
Taux d'alté- ration sur l'ensemble des variables de tous les fichiers (%)	2,5	0,24	1,1	3,85			

Note: a) Les nombres entre parenthèses indiquent le nombre de variables impliquées.

Tableau 5
Actualité des données
Temps écoulé entre la fin de l'année civile
et la date de constitution du fichier statistique
(en mois révolus)
Québec, 1975-1980

Fichier statistique	Année statistique					
	1975	1976	1977	1978	1979	1980
Naissances	31	37	30	23	11	6
Décès	30	27	16	18	6	5
Mariages	31	34	29	18	12	6
Temps moyen écoulé	30	32	25	19	9	5

complétude et d'une exactitude optimales, plus on retarde la production définitive des données d'un fichier. On ne saurait évidemment sacrifier les deux premiers objectifs pour atteindre le troisième, car la rapidité à produire des données statistiques reflétant la conjoncture démographique la plus récente possible, n'a pas de sens, si elle se fait aux dépens de la qualité.

Conclusion

L'existence de données précises et complètes est à la base de la recherche démographique. Trois éléments permettent l'évaluation de la qualité des fichiers statistiques des naissances, mariages et décès du ministère des Affaires sociales: 1) l'intégralité des fichiers statistiques, 2) l'intégrité des données et des fichiers, et 3) l'actualité des données contenues dans les différents fichiers statistiques.

Sur la base de l'année civile 1980, cette étude a permis d'estimer à près de 99,5% le degré d'intégralité des fichiers à la fin de l'année statistique. Par ailleurs on constate que le fichier statistique des décès, est celui qui subit relativement le plus d'altérations au cours de l'opération de fermeture de l'année statistique, et que celui des mariages en subit le moins. Enfin on observe que les variables les plus souvent corrigées sont les dates, les états civils et les rangs de naissance.

RÉFÉRENCES BIBLIOGRAPHIQUES

- FREDETTE, Jean-Marc, 1980, "Cueillette et traitement des données sur les naissances, les mariages, les divorces et les décès au Québec: rêves et réalité", Cahiers québécois de démographie, 9-1 (avril): 117-135.
- GABBAY, Ezra, 1975, Rapport sur le sous-enregistrement des décès, Québec, 1974, Registre de la population, ministère des Affaires sociales, Québec, collection "Données et analyses" no 3.
- GOVERNEMENT DU QUÉBEC, Loi de la protection de la santé publique, L.Q. 1972, chapitre 42, (sanctionnée le 21 décembre 1972).
- LOUP, Roland J. et Barbara J. THOMPSON, 1982, Contrôle de la qualité dans un système d'information fondé sur des dossiers médicaux, Genève, Organisation mondiale de la santé, 9 pages.
- MAHEU, Robert, 1973, "La complétude des statistiques de la naissance au Québec", Statistiques des Affaires sociales, Démographie, 1-3 (août): 72-78.
- _____, 1975, Le système de déclarations des naissances, décès et mortinaissances prévu par la Loi de la protection de la santé publique, Conférence prononcée à l'occasion d'une journée d'étude de l'Association des archivistes médicales de la province de Québec, à Montréal, (6 juin).
- ROCHON, Madeleine, 1979, Description des opérations de fermeture d'année des fichiers des naissances, mariages et décès au Registre de la population, ministère des Affaires sociales, Québec, 91 pages (non publié).
- ROY, Laurent, 1977, "Notes de présentation" dans Monographies des affaires sociales, données démographiques, Les naissances vivantes au Québec, 1950 à 1974, Révision des naissances vivantes annuelles, Québec, 1950 à 1974, Québec, Editeur officiel du Québec, janvier, p. 3-10.
- VOVAN, Kim, 1979, Processus de préparation des fichiers statistiques et de production de tableaux statistiques, Service des registres, ministère des Affaires sociales, Québec, novembre, 25 pages (non publié).