

Chapitre 1

Méthodologie

Rébecca Tremblay

Direction de la méthodologie, de la démographie et des enquêtes spéciales
Institut de la statistique du Québec

May Clarkson

Direction de l'évaluation, de la recherche et de l'innovation
Ministère de la Santé et des Services sociaux

Introduction

Ce premier chapitre vise à présenter, de façon succincte, les divers aspects méthodologiques de *l'Étude auprès des communautés culturelles 1998-1999*. Dans un premier temps, les procédures de collecte de l'information y sont abordées par une description des instruments de collecte des données, du prétest, du plan de sondage et de la taille de l'échantillon utilisés, de la collecte des données, ainsi que de l'admissibilité et de la non-réponse à l'étude. La section suivante porte sur le traitement des données et discute de leur validation, pondération et méthode d'analyse. Finalement, on décrit les normes de présentation des résultats et on conclut en exposant la portée de l'étude et certaines limites inhérentes à cette dernière.

Les lecteurs intéressés à avoir de plus amples renseignements sur la méthodologie de l'étude sont invités à consulter le cahier technique (à paraître). Par ailleurs, les divers chapitres du présent rapport décrivent plus particulièrement les aspects méthodologiques propres aux indicateurs et aux analyses traités dans leur chapitre.

1.1 Procédures de collecte

1.1.1 Instruments de collecte des données

L'Étude auprès des communautés culturelles a utilisé des instruments de collecte relativement similaires à ceux des enquêtes sociales et de santé de 1987,

de 1992-1993 et de 1998, soit : 1) une fiche d'identification portant sur l'ensemble du ménage, 2) un questionnaire administré par un interviewer (QRI) à un informateur clé dans le ménage et portant également sur l'ensemble du ménage, et 3) un questionnaire individuel autoadministré (QAA) qui s'adresse à chacune des personnes de 15 ans et plus dans le ménage. Ces instruments ont toutefois été adaptés à la population visée par l'étude.

Pour ce faire, on a pu compter sur la collaboration d'experts de la recherche auprès de communautés culturelles pour prendre connaissance de leur expérience du milieu et de leur vision des problèmes qu'ils percevaient comme spécifiques ou plus fréquents dans ces communautés; leurs commentaires ont été fort utiles dans l'élaboration des instruments. On a aussi procédé, au préalable, avec l'aide d'assistants coordonnateurs et d'assistantes coordonnatrices à l'emploi de Santé Québec et appartenant à chacun des groupes visés, à deux étapes qualitatives.

Dans un premier temps, on a réalisé des entrevues exploratoires de groupe avec des immigrants récents et des immigrants de plus longue date, appartenant à chacune des quatre populations visées par l'étude, telles que définies à la section 1.1.3. Au total, douze entrevues ont été réalisées, soit trois entrevues pour chacune des communautés visées :

1. Une entrevue avec dix hommes, immigrants récents;
2. Une entrevue avec dix femmes, immigrantes récentes;
3. Une entrevue avec cinq hommes et cinq femmes, immigrants et immigrantes de plus longue date

Ces entrevues ont permis d'explorer divers concepts, parmi lesquels on retrouve la perception de la santé et de la maladie, le processus thérapeutique et la perception des problèmes sociaux, et de dégager certaines pistes pour l'étape quantitative, notamment en ce qui concernait l'approche à adopter, les sujets spécifiques aux immigrants à aborder et ceux à éviter. Des rapports synthèses ont ensuite été produits, sous forme de monographies, pour chacune des communautés.

Dans un deuxième temps, dix entrevues individuelles à domicile ont été réalisées auprès de personnes de chacun des groupes visés, soit quarante entrevues au total, cette fois en tenant compte du ménage auquel appartenait la personne interviewée : nombre de personnes dans le ménage, type de famille, problèmes sociaux ou de santé rencontrés, recours aux services, etc. Cette étape, au cours de laquelle on a porté une attention particulière au vocabulaire employé, aux demandes de clarification et à l'ambiguïté dans l'emploi de certains termes, a permis de raffiner la démarche et les instruments pour la phase quantitative.

Les instruments, élaborés en français, ont ensuite été traduits dans les autres langues d'entrevue prévues pour l'étude, soit l'anglais, le chinois, le créole, l'arabe et l'espagnol, puis retraduits en français par des traducteurs indépendants. Lorsqu'on rencontrait des divergences de sens, les assistants coordonnateurs et assistantes coordonnatrices en discutaient avec les traducteurs, jusqu'à ce qu'on arrive à une traduction qui soit la plus fidèle possible à la version initiale en français.

Enfin, quatre nouvelles entrevues de groupe ont eu lieu, soit une pour chacune des communautés visées,

au cours desquelles les participants ont eu à répondre aux questionnaires dans leur langue respective et à commenter la formulation des questions. Cette fois, les entrevues ont été réalisées par le Groupe Léger et Léger inc.¹, qui a été mandaté pour faire la collecte des données, et ont permis d'apporter de nouvelles modifications visant à améliorer et à raffiner davantage les instruments.

1.1.2 Prétest

Quatre coordonnatrices de terrain, appartenant à chacun des groupes visés par l'étude, ont été engagées par Léger et Léger inc. Outre le fait d'avoir une bonne connaissance de la communauté qu'elles devaient représenter, elles devaient parler et lire couramment la langue vernaculaire. Le prétest final des instruments leur a été confié après qu'elles aient suivi une formation intensive d'une durée de cinq jours, assumée de façon conjointe par la firme et l'ISQ, qui portait tant sur les objectifs de l'étude, ses modalités, ses instruments et les techniques d'entrevue que sur leurs tâches de coordination. À la suite de ce prétest, on a apporté les modifications finales aux instruments.

1.1.3 Plan de sondage et taille de l'échantillon

L'Étude auprès des communautés culturelles vise la population des communautés culturelles de 18 ans et plus et leurs enfants mineurs résidant dans les régions de Montréal-Centre, de Laval ou de la Montérégie, ayant immigré entre 1988 et 1997 et appartenant à l'un des quatre groupes de pays de naissance suivants : 1) Chine, Hong Kong, Taïwan, Macao; 2) Haïti; 3) pays arabophones d'Afrique du Nord et du Moyen-Orient; 4) pays hispanophones des Amériques. Des renseignements sur le choix de ces quatre groupes ont été précédemment donnés dans l'introduction générale. Lors de la préparation de l'étude, la base de sondage indiquait que la population de 18 ans et plus de ces groupes de pays de naissance couvrait 55,6 % de la population immigrante de 18 ans et plus de 1986-1992.

1. Devenu depuis Léger Marketing inc.

Un échantillon de 3 000 ménages appartenant à cette population a été tiré de façon aléatoire à partir des fichiers du ministère des Relations avec les citoyens et de l'Immigration et leur adresse la plus récente a été obtenue par l'intermédiaire de la Régie de l'assurance maladie du Québec. Un ménage pouvait faire partie de l'échantillon à plus d'une reprise, étant donné que la base de sondage était en fait constituée d'individus, considérés points de contact avec les ménages. Il pouvait ainsi arriver que deux individus ou plus soient un point de contact pour le même ménage, ce qui est arrivé à seulement cinq reprises.

Avant de procéder à la sélection de l'échantillon, la population de 18 ans et plus a été découpée en vingt-quatre parties, appelées strates, formées par le croisement de trois variables : la région d'origine (Chine, Haïti, Maghreb et Moyen-Orient, pays hispanophones), la période d'immigration (1988-1992, 1993-1997) et la catégorie d'immigrants (indépendant, famille, réfugié). L'échantillon a été tiré aléatoirement à partir de chacune des strates, comme présenté au

tableau 1.1, et les données sur l'ensemble du ménage ont été recueillies. Le plan de sondage est donc stratifié par grappes où une grappe représente un ménage. Ce type de plan entraîne une concentration de l'échantillon au niveau du ménage conduisant à une perte de précision par rapport à un plan aléatoire simple mais produisant un échantillon de plus grande taille pour un coût comparable de collecte.

L'échantillon des 3 000 ménages, dont 750 par communauté, est réparti de façon non proportionnelle à la population selon le groupe de pays de naissance afin de réaliser l'objectif d'analyse par région d'origine (tableau 1.2). Cela entraîne toutefois une perte de précision des statistiques sur l'ensemble de ces groupes par rapport à un plan aléatoire simple.

Tableau 1.1

Répartition de l'échantillon selon la région d'origine, la période d'immigration et la catégorie d'immigrants, communautés culturelles 1998-1999

	Période	Catégorie d'immigrants			Total
		Indépendant	Famille	Réfugié	
Chine	1988-1992	329	81	10	420
	1993-1997	202	118	10	330
Haïti	1988-1992	47	340	23	410
	1993-1997	17	279	44	340
Maghreb et Moyen-Orient	1988-1992	321	81	58	460
	1993-1997	150	100	40	290
Pays hispanophones	1988-1992	95	119	241	455
	1993-1997	47	141	107	295
<i>Total</i>	<i>1988-1992</i>	<i>792</i>	<i>621</i>	<i>332</i>	<i>1 745</i>
<i>Total</i>	<i>1993-1997</i>	<i>416</i>	<i>638</i>	<i>201</i>	<i>1 255</i>
Total	1988-1997	1 208	1 259	533	3 000

Source : Institut de la statistique du Québec, *Étude auprès des communautés culturelles 1998-1999*.

Tableau 1.2

Distribution de la population et de l'échantillon selon la région d'origine, communautés culturelles 1998-1999

	Population	Échantillon
	%	
Chine	14,2	25,0
Haïti	14,7	25,0
Maghreb et Moyen-Orient	49,1	25,0
Pays hispanophones	22,0	25,0
Total	100,0	100,0

Source : Institut de la statistique du Québec, *Étude auprès des communautés culturelles 1998-1999*.

Pour chaque communauté, la répartition de l'échantillon est proportionnelle à la racine carrée de la répartition de sa population selon la période d'immigration. Cela a pour effet d'améliorer la précision des statistiques calculées sur la période d'immigration 1993-1997 tout en conservant une bonne précision sur les statistiques calculées sur les deux périodes ensemble, et ce, toujours en comparaison avec une répartition proportionnelle de l'échantillon. Finalement, pour chaque communauté et période d'immigration, l'échantillon est réparti de façon proportionnelle à la population selon la catégorie d'immigrants.

1.1.4 Collecte des données

La collecte des données s'est déroulée de novembre 1998 à la fin août 1999 mais devait se terminer beaucoup plus tôt. Plusieurs problèmes sont apparus en cours de collecte et le taux de réponse (voir la section « portée et limites des données ») était très faible. Il a donc été décidé, lors d'une rencontre avec le comité scientifique de Santé Québec qui a eu lieu en mars 1999, de modifier la méthodologie et de poursuivre la collecte avec une seconde équipe d'interviewers.

À l'origine, cette collecte a été confiée à des interviewers appartenant à chacune des communautés, supervisés par les quatre coordonnatrices qui ont effectué le prétest. Le recrutement de trente-deux interviewers, au nombre de huit pour chaque communauté, a été effectué avec l'aide des

coordonnatrices, qui ont d'ailleurs assuré leur formation. Cette dernière, d'une durée de cinq jours, se déroulait concurremment en chinois (cantonais ou mandarin), en créole, en arabe et en espagnol, suivant les modalités indiquées dans le cahier de formation préparé par Santé Québec, le BSQ et la firme. Ce cahier étant rédigé en français, on s'était assuré au préalable que chaque interviewer pouvait au moins lire le français. Le directeur de terrain du Groupe Léger et Léger inc. et la coordonnatrice de l'étude assuraient une présence constante lors de la semaine de formation afin de répondre, le cas échéant, aux questions soulevées par les coordonnatrices ou les interviewers. Il était cependant difficile de contrôler la qualité de la formation donnée à ces interviewers en raison des langues de transmission qui ont été utilisées.

Les difficultés survenues en cours de collecte proviennent, d'une part, du fait que les interviewers étaient généralement jeunes et inexpérimentés et qu'ils avaient sans doute sous-estimé les difficultés de leur tâche, et, d'autre part, que les personnes rencontrées pouvaient hésiter à se confier, pour des raisons variables selon les cultures (par exemple, le jeune âge de l'interviewer ou la crainte d'un bris de confidentialité). Combiné à ces problèmes, le faible taux de réponse en cours de collecte a conduit à la décision d'embaucher des interviewers chevronnés, qui venaient de terminer leur affectation à *l'Enquête sociale et de santé 1998*. Ces nouveaux interviewers ont pris la relève de la première équipe, accompagnés au besoin par des interprètes, après avoir suivi au préalable la même formation intensive de cinq jours.

Dans chaque ménage échantillonné, une entrevue était effectuée auprès d'un informateur clé, âgé de 18 ans ou plus, appartenant au groupe visé par l'étude. Il s'agissait, dans la mesure du possible, de la personne la mieux informée sur l'état de santé des membres du ménage et sur leur consommation de soins et de médicaments. Cette personne devait répondre au QRI pour chacun des membres du ménage et pour elle-même. Lorsque personne n'était présent lors de la première visite, l'interviewer devait

effectuer jusqu'à six tentatives pour établir le contact, en variant l'heure et le jour de ces tentatives.

Lorsque le QRI était rempli, l'informateur clé devait également remplir un QAA, après quoi l'interviewer lui remettait une compensation monétaire de 15 \$. Les autres personnes de 15 ans et plus dans le ménage étaient également invitées à remplir un QAA, qu'elles pouvaient immédiatement remettre à l'interviewer dans une enveloppe scellée ou expédier plus tard à la firme dans une enveloppe préaffranchie.

La première équipe d'interviewers acheminait les QRI aux coordonnatrices, qui en assuraient la vérification avant de les transmettre à la firme tandis que la seconde équipe transmettait les QRI directement à la firme, qui en assurait alors la vérification. Quant aux QAA recueillis sur place par les interviewers, ils étaient acheminés par messagerie à la firme de sondage.

1.1.5 Admissibilité et non-réponse

La taille de l'échantillon de l'*Étude auprès des communautés culturelles* est constituée du nombre total de ménages sélectionnés pour l'étude, à partir desquels la collecte des données s'effectue. Toutefois, lors de la collecte, certains des ménages échantillonnés s'avèrent inadmissibles à l'étude selon des critères déterminés *a priori* portant sur l'individu servant de point de contact du ménage, tandis que d'autres ne répondent pas à l'étude. Il y a également une partie des ménages échantillonnés dont l'admissibilité est incertaine et qui nécessitent ainsi une attention particulière.

- Admissibilité à l'étude

Il y a 3 000 ménages qui ont été sélectionnés pour participer à l'étude et certains ménages se sont avérés inadmissibles à cette dernière, de façon définitive ou potentielle (tableau 1.3). Les 211 ménages déclarés définitivement non admissibles le sont d'après les critères fixés avant la collecte tandis que les 1 018 ménages jugés potentiellement admissibles ont été soumis à une étude approfondie de leur

admissibilité d'après leur individu servant de point de contact, et ce, à l'aide des données de l'*Étude auprès des communautés culturelles* et de données provenant de la Régie de l'assurance maladie du Québec qui portaient sur l'admissibilité de ces individus au régime d'assurance maladie et sur leurs consultations médicales. Deux recommandations ont alors été faites pour traiter ces ménages potentiellement admissibles. La première a été d'estimer à 57,4 % la proportion de ménages inadmissibles à l'étude parmi ceux ayant déménagé on ne sait où (mauvaise adresse), sans toutefois pouvoir clairement les identifier. La seconde recommandation a été d'exclure des ménages admissibles ceux pour lesquels aucune adresse n'était disponible (aucune adresse). Cela conduit ainsi à une estimation de 638 ménages inadmissibles supplémentaires². Les 849 ménages estimés inadmissibles sont donc composés en grande partie de ménages ayant une mauvaise adresse (60 %), n'ayant aucune adresse (15 %) ou n'habitant plus la région couverte (15 %).

Ainsi, il y a 2 151 ménages qui sont estimés admissibles au QRI. Ils représentent la population visée qui est évaluée approximativement à 107 444 individus (tableau 1.4). Cette estimation provient de l'ajustement à l'admissibilité de la population de la base de sondage. Par la suite, on estime à 3 019 personnes de 15 ans et plus le nombre d'individus admissibles au QAA, correspondant à une population visée de 71 982 individus âgés de 15 ans et plus. Les estimations de population sont données à titre indicatif seulement car la population à l'étude est fort mobile et la qualité des estimations présentées est variable selon la communauté en raison de l'ajustement pour l'inadmissibilité.

2. Voir l'« Étude de l'admissibilité à l'enquête » de Plante, Courtemanche et Tremblay (document interne à la DMDES de l'ISQ). Le cahier technique de l'étude (à paraître) donnera également de plus amples détails.

Tableau 1.3

Répartition des ménages échantillonnés selon les résultats de la collecte pour le QRI, communautés culturelles 1998-1999

	Résultats observés			Résultats pondérés	
	Nombre de ménages	non pondéré	Estimation de ménage admissibles		
	n	%	n	%	
Ménages définitivement non admissibles					
Pays non couvert ou aucun nom <i>a priori</i>	58	1,9	
N'habite plus la région couverte	129	4,3	
Non originaire de l'un des pays couverts ou a immigré au Québec avant 1988	24	0,8	
<i>Sous-total</i>	<i>211</i>	<i>7,0</i>	
Ménages potentiellement admissibles					
Mauvaise adresse	892	29,7	380	17,8	
Aucune adresse	126	4,2	0	0,0	
<i>Sous-total</i>	<i>1 018</i>	<i>33,9</i>	<i>380</i>	<i>17,8</i>	
Ménages définitivement admissibles					
Entrevue complétée	1 161	38,7	1 161	53,6	
Refus	449	15,0	449	21,0	
Impossibilité de répondre	161	5,4	161	7,6	
<i>Sous-total</i>	<i>1 771</i>	<i>59,0</i>	<i>1 771</i>	<i>82,2</i>	
Total	3 000	100,0	2 151	100,0	

Source : Institut de la statistique du Québec, *Étude auprès des communautés culturelles 1998-1999*.

- Non-réponse totale

Parmi les 2 151 ménages estimés admissibles à l'étude, 1 161 ménages ont répondu au QRI. Les principales raisons de non-réponse au QRI sont le refus du ménage et l'obtention d'une mauvaise adresse (45 % et 38 % des non-répondants) (tableau 1.3). Les ménages répondants au QRI abritent au total 4 162 personnes visées par l'étude, dont 3 019 personnes de 15 ans et plus qui étaient admissibles au QAA; seulement 1 871 personnes ont rempli ce second questionnaire (tableau 1.4).

Le taux de réponse à un instrument de collecte est défini comme étant le rapport entre le nombre d'unités répondantes et le nombre d'unités admissibles à cet instrument. Dans la présente étude, comme toutes les unités n'ont pas la même probabilité d'être choisies, il est préférable de considérer un taux de réponse établi sur la base des données pondérées par l'inverse de cette probabilité. Ce taux pondéré est ainsi comparable à tout autre taux de réponse pondéré d'une enquête sur la même population, indépendamment de la répartition de l'échantillon.

Tableau 1.4

Nombre de répondants pour le QRI et pour le QAA selon la région d'origine, communautés culturelles 1998-1999

	Population estimée admissible au QRI	Nombre de répondants au QRI	Population estimée admissible au QAA	Nombre de répondants au QAA
Chine	12 606	950	10 012	543
Haïti	17 666	1 068	10 896	379
Maghreb et Moyen-Orient	50 024	960	34 017	405
Pays hispanophones	27 148	1 184	17 057	544
Total	107 444	4 162	71 982	1 871

Source : Institut de la statistique du Québec, *Étude auprès des communautés culturelles 1998-1999*.

Pour le QRI, le taux de réponse pondéré est établi sur la base des ménages échantillonnés admissibles et s'élève à 53,6 %. Les personnes de 15 ans et plus ont ensuite répondu au QAA dans une proportion de 63,5 %, cette proportion étant établie sur la base des données pondérées. Cependant, le taux de réponse au QAA doit tenir compte du fait que les questionnaires sont administrés en cascade et il se définit par le produit du taux de réponse au QRI et de cette proportion (63,5 %). Le taux de réponse au QAA ainsi calculé est de 34,0 % (tableau 1.5). Par ailleurs, la proportion pondérée de répondants au QAA parmi les individus admissibles et le taux de réponse pondéré au QRI varient selon la communauté.

Tableau 1.5

Proportion pondérée de répondants et taux de réponse pondéré aux questionnaires selon la région d'origine, communautés culturelles 1998-1999

	Proportion pondérée de répondants ¹ %	Taux de réponse pondéré
QRI	...	53,6
Chine	...	47,6
Haïti	...	56,2
Maghreb et Moyen-Orient	...	52,0
Pays hispanophones	...	58,8
QAA	63,5	34,0
Chine	70,9	33,7
Haïti	56,2	31,6
Maghreb et Moyen-Orient	60,1	31,3
Pays hispanophones	70,7	41,6

1. Parmi les individus admissibles et ayant répondu au QRI.

Source : Institut de la statistique du Québec, *Étude auprès des communautés culturelles 1998-1999*.

Il existe également d'autres sources de variabilité du taux de réponse au QRI et de la proportion pondérée de répondants au QAA par rapport aux personnes admissibles; ces sources peuvent être différentes, ou encore avoir moins d'importance, selon la communauté étudiée. Par exemple, l'âge est une source de variabilité particulièrement importante pour le QRI chez les immigrants récents d'origine chinoise : le taux de réponse à ce questionnaire est moins grand

pour les ménages dont la personne échantillonnée est âgée de 65 ans et plus (données non présentées). Toutefois, dans la population originaire du Maghreb ou du Moyen-Orient, c'est le statut d'immigration qui est associé au taux de réponse au QRI : les ménages dont l'individu échantillonné a un statut d'immigrant indépendant sont davantage représentés parmi les non-répondants (données non présentées). En ce qui concerne le QAA, le répondant principal de chaque ménage a répondu en plus grande proportion que les autres répondants du ménage, et ce, pour l'ensemble des communautés de même que pour chacune d'elles (données non présentées). Certains ajustements sont faits pour minimiser l'effet potentiellement néfaste de la non-réponse totale sur les estimations produites (voir la section 1.2.2).

▪ Non-réponse partielle

En plus de la non-réponse totale, il faut que les estimations tiennent compte de la non-réponse à chacune des questions, soit la non-réponse partielle. L'ampleur de ce phénomène est mesurée par le taux de non-réponse partielle, qui est défini comme étant le rapport entre le nombre pondéré de non-répondants à une question donnée et le nombre pondéré de personnes devant y répondre. Les taux de non-réponse partielle au QRI sont généralement faibles étant donné la présence de l'interviewer, mais ils sont plus importants au QAA. Toutefois, les estimations produites dans cette étude ne comportent pas de correction pour la non-réponse partielle.

Puisqu'il arrive que la non-réponse partielle à certaines questions atteigne des seuils critiques, on a effectué une analyse des caractéristiques des non-répondants lorsque le taux de non-réponse partielle était supérieur à 10 %. Les non-répondants, par exemple, peuvent être caractérisés par leur communauté, leur classe d'âge, leur sexe ou leur statut d'activité au cours d'une période de 12 mois. La section « Aspects méthodologiques » de chaque chapitre présente les taux de non-réponse partielle des variables principales lorsqu'ils sont supérieurs à 10 %, les caractéristiques des non-répondants et leur effet éventuel sur les résultats.

1.2 Traitement des données

1.2.1 Validation

Avant toute codification, une première vérification de logique du questionnaire a été effectuée par le Groupe Léger et Léger inc., et ce, pour chacun des instruments (fiche d'identification, QRI et QAA). Parallèlement à cette vérification, la firme effectuait un contrôle du travail de chacun des interviewers par la réadministration téléphonique, auprès de 10 % des informateurs clés des ménages sélectionnés au hasard, d'une partie du QRI, afin de vérifier la qualité de l'entrevue et la fiabilité des renseignements recueillis. En ce qui concerne les QAA, on vérifiait également la qualité de l'autocodage avant qu'ils soient acheminés à la saisie.

Pour ce qui est de la codification des questions ouvertes du QRI, la firme a eu recours à des codeurs spécialisés, les mêmes types de codeurs que dans le cas de *l'Enquête sociale et de santé 1998*: pharmaciens pour la codification des médicaments, archivistes médicaux pour les problèmes de santé et codeurs d'expérience pour les professions et secteurs d'activité.

Chacun des instruments a fait l'objet d'une double saisie informatique. Les fichiers des trois instruments ont ensuite été acheminés à l'ISQ qui a effectué diverses vérifications portant sur la gestion de l'échantillon et de la collecte telles que la concordance entre la liste des ménages fournie par la firme et l'échantillon initial, la vérification des codes d'inadmissibilité ou de non-réponse pour les ménages ou les individus, etc. On a également contrôlé le travail de validation effectué par la firme. Enfin, l'ISQ a procédé à des validations sur la cohérence des données, par le biais d'une série de croisements complexes entre les questions, ainsi qu'à une validation interinstruments.

1.2.2 Pondération

Les données obtenues ont été ensuite pondérées. Cela consiste à attribuer à chaque répondant un

poids, qui correspond au nombre de personnes qu'il « représente » dans la population. Pour chaque instrument de collecte de l'étude, on attribue donc à chaque répondant des facteurs de pondération différents pour le QRI et pour le QAA.

Le calcul du poids du répondant à un questionnaire tient compte de deux éléments : la probabilité de sélection et la non-réponse totale. La pondération est établie de façon consécutive : on définit d'abord la pondération pour le QRI, puis celle pour le QAA.

La pondération pour le QRI tient compte de la probabilité de sélection et de la non-réponse totale au QRI, intégrées en une seule étape appelée probabilité de sélection et de réponse. Dans les enquêtes usuelles, ces deux étapes se suivent habituellement dans l'établissement de la pondération. La particularité de cette étude vient du fait que le ménage est indirectement échantillonné par la sélection d'un point de contact du ménage, qui est l'une ou l'autre des personnes visées du ménage, et que l'information contenue dans la base de sondage porte sur ces points de contact, et non sur les ménages. La pondération ménage est déterminée par l'inverse de la probabilité de sélection et de réponse du ménage, qui est une fonction des probabilités de sélection et de réponse de chaque individu du ménage de même que de leurs probabilités conjointes de sélection et de réponse. On attribue ensuite la pondération ménage à chaque individu visé du ménage, leur fixant ainsi une pondération individuelle pour cet instrument.

L'utilisation de la probabilité de sélection permet de considérer la non-proportionnalité de l'échantillon par rapport à la distribution de la population (tableau 1.2). Ainsi, les unités appartenant à des strates sous-représentées ont des poids initiaux plus élevés que celles provenant de strates surreprésentées. La variable « période d'immigration », nécessaire au calcul de la probabilité de sélection des individus mais non fournie pour tous dans la base de sondage (seulement pour les individus points de contact) ni par les questionnaires de l'étude, a dû être imputée pour ce faire. Il est

important de noter que cette variable n'est pas nécessairement équivalente à la variable « période d'établissement », de par leur définition et leurs valeurs respectives; la période d'établissement peut être construite à partir de l'année d'établissement demandée dans le QRI.

L'ajustement pour la non-réponse totale permet de minimiser les biais dans les estimations, qui sont causés par une représentation inadéquate de la population lorsque les non-répondants ont des caractéristiques différentes de celles des répondants. Les résultats de la collecte sur les ménages échantillonnés ont permis d'analyser la non-réponse totale au QRI à l'aide de modèles de régression logistique. Les classes formées par le croisement de variables explicatives au seuil de 5 % et les taux de réponse ainsi obtenus de ces modèles ont permis de fixer une probabilité de réponse à chaque individu.

La pondération pour le QAA se base sur la pondération pour le QRI. Ces poids sont corrigés pour la non-réponse totale au QAA grâce à des classes d'ajustement définies à l'aide d'une méthode de modélisation par segmentation de la probabilité de réponse. Pour ce faire, on a employé l'algorithme CHAID (Chi-Square Automatic Interaction Detection/ détection automatique d'interactions du khi carré) mis au point par Kass (1980). Les variables expliquant le mieux la réponse de chaque communauté pouvant différer, la construction des classes d'ajustement pour la non-réponse totale au QAA a été faite séparément pour ces quatre populations. La correction consiste à multiplier le poids du QRI par l'inverse de la proportion pondérée de répondants au QAA parmi les individus admissibles, observée au sein de chaque classe d'ajustement créée. Les variables telles que le type de répondant (informateur clé ou non), si c'est la personne échantillonnée (point de contact du ménage) ou non, la classe d'âge, le sexe, la catégorie d'immigrants et la période d'immigration ont été liées au fait de répondre ou non dans l'une ou l'autre des communautés. Finalement, les poids sont ajustés à l'estimation de la population adulte visée selon la communauté, l'âge et le sexe, à l'aide d'un simple

facteur d'ajustement comme multiplicateur pour les poids de tous les répondants appartenant à un même groupe d'âge et de sexe à l'intérieur d'une communauté donnée. L'ajustement selon le sexe n'a cependant pas été fait chez les 65 ans et plus en raison des petits effectifs.

À toutes les étapes de pondération, des précautions ont été prises relativement aux poids extrêmes afin de ramener le plus possible les valeurs jugées aberrantes ou extrêmes à l'intérieur de limites raisonnables.

1.2.3 Méthodes d'analyse

À l'instar de plusieurs enquêtes de santé, la présente étude a sollicité la participation de nombreux professionnels et chercheurs du réseau de la santé, des services sociaux et des universités. On a mis sur pied des groupes d'analyse, composés en général d'une à trois personnes choisies pour leur expertise et leur intérêt pour un sujet traité dans l'étude. Ils ont proposé un plan d'analyse des données portant sur leur thème et formulé des demandes de tableaux. Ces derniers ont été produits par l'ISQ et analysés par les groupes. Le tout a ensuite été vérifié par l'ISQ.

Les analyses proposées dans le présent rapport sont descriptives et elles sont essentiellement univariées et bivariées. Cette orientation a été retenue pour que les analyses demeurent simples et plus facilement compréhensibles tout en espérant de rendre les résultats accessibles dans des délais raisonnables aux planificateurs, aux décideurs et aux chercheurs. Quelques analyses multivariées ont aussi été réalisées. L'ensemble des analyses de ce rapport portent principalement sur des variables catégoriques quoique certaines variables continues ont également été étudiées. Par ailleurs, le recours aux données pondérées rend possible l'inférence à la population visée.

Dans la majorité des enquêtes statistiques, les estimations produites comportent des erreurs d'échantillonnage, c'est-à-dire qu'elles sont dues au fait que seulement une partie de la population visée

est interrogée. La complexité du plan de sondage utilisé pour produire l'échantillon de l'*Étude auprès des communautés culturelles* influence grandement la précision des résultats; une mesure de la précision de chaque estimation est donc nécessaire à l'inférence, à l'analyse et à l'interprétation juste des résultats.

Le coefficient de variation permet de mesurer la précision relative d'une estimation. On l'obtient en divisant l'erreur-type de l'estimation par l'estimation elle-même. Il a été calculé pour chaque estimation de proportion effectuée. Par ailleurs, une autre mesure de précision, l'effet de plan, renseigne sur la perte ou le gain de précision associés à une estimation attribuable au plan de sondage par rapport à la précision qui aurait été obtenue avec un échantillon aléatoire simple de même taille. Étant donné le très grand nombre d'estimations produites dans ce rapport, des effets de plan moyens ont été calculés sur des proportions pour quelques catégories de variables homogènes par rapport à la variance. De plus, puisque les estimations peuvent porter sur chaque communauté tout comme sur l'ensemble des communautés, ces effets de plan moyens ont été évalués pour chacune de ces populations.

En ce qui concerne le QRI, pour les estimations portant sur les personnes, quatre catégories de variables ont été nécessaires et les quatre effets de plan moyens correspondants ont été produits, par communauté et globalement. L'effet de plan moyen principal associé à la catégorie de variables regroupant la majorité des phénomènes étudiés s'élève à 2,0 pour l'ensemble des communautés. Cela signifie que les 4 162 répondants au QRI permettent d'établir des estimations ayant une précision équivalente à celle qu'un échantillon aléatoire simple de 2 081 personnes aurait fournie. Pour les estimations provenant du QAA et portant sur les personnes, seulement deux catégories de variables ont suffi et deux effets de plan moyens ont été déterminés par communauté et globalement. Pour la majorité des caractéristiques étudiées, un effet de plan moyen global de 2,1 est obtenu. Par ailleurs, des effets de plan spécifiques supplémentaires ont aussi

été calculés pour certains indices : 1) pour les estimations d'indices du QRI utilisant les ménages ou les médicaments comme unités d'analyse au lieu des individus, 2) lorsqu'un indice du QRI ou du QAA utilise dans sa construction des variables ayant des effets de plan différents (ex. : une variable avec l'effet de plan général et une variable avec un effet de plan particulier).

Des intervalles de confiance ont également servi à l'analyse des données. On les a calculés en tenant compte du plan de sondage de l'*Étude auprès des communautés culturelles*. Ces intervalles ne sont pas indiqués dans ce rapport, à l'exception de ceux calculés lors des analyses multivariées et de ceux pour les estimations de moyennes du chapitre 11; on leur a préféré le coefficient de variation pour indiquer la précision des estimations de proportions.

En ce qui a trait aux analyses bivariées, des tests globaux du khi carré ont été effectués, soit sur l'association de deux variables dans une population ou sous-population donnée, soit sur un contraste linéaire de paramètres. Lorsqu'un test global était significatif à un seuil de 5 %, on pouvait poursuivre en effectuant un test de comparaison de proportions ou de moyennes, également à un seuil de 5 %. Dans tous les cas, on a procédé à un ajustement du test usuel pour tenir compte de la complexité du plan de sondage³. Notez par ailleurs qu'aucun test global sur l'association d'une variable d'intérêt avec la communauté n'a été fait puisqu'il a été décidé de ne pas comparer les communautés entre elles en raison de leur hétérogénéité d'origines et de cultures. Tout au plus, on a comparé une communauté particulière avec les autres communautés prises ensemble.

Lorsque cela était pertinent, une analyse de régression multiple, logistique ou linéaire, a pu être entreprise de façon exploratoire; son but n'était cependant pas de modéliser parfaitement la variable dépendante d'intérêt. Ces régressions ont été réalisées en fonction des résultats provenant des

3. Le cahier technique de l'étude (à paraître) donnera de plus amples détails sur les tests utilisés et leurs corrections.

analyses bivariées : les associations observées avec la variable d'intérêt, en l'occurrence la variable dépendante du modèle, ont permis de choisir des variables indépendantes à conserver pour l'analyse en plus de certaines variables de contrôle. L'impact de la non-réponse partielle globale de chaque modèle n'a pas été étudié mais les taux de non-réponse partielle des variables en jeu ont été vérifiés afin de la contrôler le plus possible. Généralement, les modèles finaux sont constitués d'interactions doubles ou de variables significatives à un seuil de 5 % en plus de variables confondantes. Lorsqu'un test global sur un effet (variable ou interaction) était significatif à ce seuil, l'analyse pouvait alors se poursuivre par l'étude des coefficients de régression estimés. Pour la régression logistique, l'estimation de rapports de cotes a plutôt été privilégiée pour effectuer l'interprétation.

Afin d'effectuer les comparaisons de la population immigrante récente visée par l'*Étude auprès des communautés culturelles* avec la population du Québec en général, les données de l'*Enquête sociale et de santé 1998* (Daveluy et autres, 2000) ont été standardisées selon l'âge et le sexe. Sans cela, étant donné que cette population est notamment différente sur le plan de sa structure d'âge, la population du Québec en général étant plus âgée que la population immigrante visée (tableau 2.9, chapitre 2), aucune comparaison de statistiques portant sur l'ensemble de la population n'aurait été intéressante, ce qui aurait limité grandement les comparaisons⁴. La méthode de standardisation directe (Bogue, Arriaga et Anderton, 1993) a été privilégiée et la population de référence est la population immigrante visée. Plus précisément, cinq populations de référence sont disponibles pour les analyses, soit celle de l'ensemble des communautés culturelles ou encore celle d'une communauté en particulier lors d'une comparaison avec une seule d'entre elles (populations chinoise, haïtienne, arabophone ou hispanophone). Les tests de comparaison ont été effectués en tenant compte des plans de sondage complexes de ces enquêtes. Ils ont

4. Dans ce cas, seules des comparaisons par groupe d'âge auraient été alors acceptables mais les petits effectifs pour certains groupes d'âge restreignaient les analyses.

tous été réalisés à un seuil de 5 %, sans toutefois être ajustés pour les comparaisons multiples.

1.3 Présentation des résultats

Cette section présente quelques indications générales s'appliquant à l'ensemble des résultats concernant notamment leur formulation générale et la présentation des estimations.

Toutes les estimations présentées dans les tableaux de ce rapport portent sur la population des communautés culturelles de 18 ans et plus et leurs enfants mineurs résidant dans les régions de Montréal-Centre, de Laval ou de la Montérégie, ayant immigré entre 1988 et 1997 et appartenant à l'un des quatre groupes de pays de naissance visés, sauf indication contraire. Ainsi, malgré que l'ensemble des personnes des ménages soient interrogées pour l'étude, seules les données des personnes visées sont employées pour effectuer les estimations des caractéristiques individuelles de la population visée, comme l'estimation de la proportion de buveurs actuels dans la population de 15 ans et plus. Les données des individus non visés des ménages ont été utilisées seulement dans le but de créer des variables portant sur le ménage.

Les statistiques fournies dans ce rapport sont établies à partir d'un échantillon et comportent donc des erreurs d'échantillonnage. La présentation des résultats en rend généralement compte en utilisant des expressions montrant qu'il ne s'agit pas de valeurs exactes. Dans les tableaux, une mesure de précision, le coefficient de variation, a été calculée pour chaque proportion ou moyenne. Parce qu'elles sont suffisamment précises, les estimations dont le coefficient de variation est inférieur ou égal à 15 % sont présentées sans commentaire; celles dont le coefficient de variation se situe entre 15 % et 25 % sont marquées d'un astérisque (*) pour montrer que leur précision est passable et qu'elles doivent être interprétées avec prudence. Les estimations dont le coefficient de variation est supérieur à 25 % sont marquées d'un double astérisque (**) pour signaler leur faible précision et indiquer qu'elles doivent être

utilisées avec circonspection; elles ne sont fournies qu'à titre indicatif. Lorsque le coefficient de variation n'était pas employé, un intervalle de confiance à 95 % a plutôt été présenté.

L'interprétation des coefficients de régression est faite uniquement de façon qualitative. Les paramètres de régression linéaire et les rapports de cotes sont tout de même présentés dans les tableaux, accompagnés d'intervalles de confiance à 95 % comme mesure de précision. En aucun cas les rapports de cotes ne doivent être interprétés sous l'angle de rapport de prévalences (connu aussi sous l'appellation de risque relatif) car ce type d'interprétation n'est valable que lorsque la proportion à estimer représente un événement rare.

Les estimations présentées ont été arrondies à l'unité quand elles sont mentionnées dans le texte, et à une décimale dans les tableaux, à l'exception de celles inférieures à 10 % pour lesquelles une décimale a aussi été conservée dans le texte. En raison de l'arrondissement, la somme des proportions de certains tableaux peut être légèrement supérieure ou inférieure à 100 %.

En général, seuls les résultats significatifs au seuil de 5 % sont mentionnés dans le texte. Il peut par ailleurs arriver que deux estimations semblent différentes ne le soient pas d'un point de vue statistique, à cause notamment du petit nombre d'individus sur lequel est basée l'estimation. Dans ce cas, le texte indique qu'il n'y a pas de différence statistiquement significative ou que l'étude ne permet pas de noter de différence entre ces estimations. Dans certains cas exceptionnels, des résultats non significatifs au seuil de 5 % mais situés près de cette limite peuvent être signalés s'ils présentent un intérêt particulier en matière de santé; ils sont alors exprimés sous forme de « tendances ».

Lorsqu'un indicateur est « mesuré » sur une période de référence donnée, celle-ci est précisée dans les aspects méthodologiques du chapitre où l'indicateur est analysé. Par exemple, la question 18 du QRI est

énoncée comme suit : « Au cours des 2 dernières semaines, quelqu'un du foyer s'est-il adressé à un médecin généraliste? » La période de référence est les deux semaines précédentes. Cependant, puisque la collecte des données de *l'Étude auprès des communautés culturelles* a duré environ sept mois, les répondants ne renvoient pas tous aux mêmes deux semaines. Ainsi, les estimations n'ont pas été présentées en désignant les deux semaines ayant précédé l'étude mais plutôt en faisant état de résultats pour une période de deux semaines.

Les estimations standardisées portant sur la population québécoise peuvent être différentes de celles qui sont publiées dans le rapport de *l'Enquête sociale et de santé 1998* (Daveluy et autres, 2000). En effet, les estimations standardisées québécoises dont il est question dans ce rapport ne renseignent pas exactement sur les caractéristiques de la population québécoise. Cependant, elles permettent une comparaison adéquate avec les données des communautés culturelles étudiées ici. Les estimations standardisées québécoises publiées dans ce rapport ont comme population de référence l'ensemble des communautés culturelles et ne peuvent être directement comparées qu'avec les estimations pour l'ensemble des communautés. Pour des fins d'analyse comparative, on a également calculé des estimations québécoises standardisées ayant comme population de référence chacune des communautés visées. Ces estimations ne sont pas incluses dans les tableaux; les résultats des analyses sur ces dernières peuvent cependant être mentionnés dans le texte.

1.4 Portée et limites de l'étude

Tout a été mis en place pour assurer la qualité et la représentativité de *l'Étude auprès des communautés culturelles*. Plusieurs entrevues, de groupe de même qu'individuelles, ont été réalisées pour mettre cette étude sur pied et la raffiner à diverses étapes. Les instruments ont été élaborés à partir de ceux de l'ESS 1998, enquête bien rodée qui en était à sa troisième répétition, en prenant soin de les adapter aux expériences et aux problèmes plus spécifiques

aux populations immigrantes à l'étude grâce, entre autres, à la collaboration d'intervenants du milieu et d'assistants coordonnateurs et assistantes coordonnatrices qui appartiennent aux communautés visées. Ces instruments ont été par la suite traduits dans les diverses langues prévues pour l'étude dans le but de contrecarrer les barrières linguistiques possibles.

Paradoxalement, c'est lorsqu'on a confié le travail sur le terrain à des interviewers issus de chacune des communautés à l'étude qu'on a rencontré davantage de problèmes. Les taux de réponse demeuraient assez faibles à ce stade de la collecte des données et les interviewers et les coordonnatrices de terrain étaient découragés et démotivés; il a donc fallu compléter la collecte en ayant recours à des interviewers plus âgés et très expérimentés, assistés au besoin d'un interprète, plus particulièrement pour récupérer les refus. Cette nouvelle approche a permis d'obtenir des taux de réponse tout de même acceptables, considérant ceux habituellement obtenus dans des enquêtes portant sur cette population difficile à joindre et les difficultés éprouvées lors de la première partie de la collecte.

Du point de vue statistique, l'étude utilise un échantillon de taille relativement considérable permettant de décrire adéquatement l'ensemble de la population visée. Cependant, une fois ventilé par communauté, il limite parfois les analyses à l'intérieur de ces sous-groupes. Sur le plan de la réponse totale, la collecte a permis d'atteindre un taux de réponse au QRI de 53,6 % et une proportion de 63,5 % de répondants au QAA parmi les individus de 15 ans et plus répondants au QRI. Ainsi, quelque 1 200 ménages ont été interrogés, ce qui correspond à plus de 4 000 immigrants récents de tous âges pour les données provenant du QRI, et à environ 2 000 immigrants récents pour les données provenant du QAA. Par ailleurs, la présence de l'interviewer pour le QRI a permis d'éviter le plus possible la non-réponse partielle; cette dernière a été plus importante pour le QAA. La non-réponse partielle a cependant été examinée pour chaque variable étudiée dans le présent rapport, et les chapitres fournissent des

indications sur le sens des biais potentiels à considérer dans l'analyse.

Une attention particulière a été accordée aux procédures inférentielles utilisées dans *l'Étude auprès des communautés culturelles*. Premièrement, des pondérations ont été effectuées pour chaque instrument de collecte de façon à assurer la fiabilité de l'inférence à la population visée et à minimiser les biais potentiels associés à la non-réponse totale; ces pondérations sont utilisées dans les analyses des données de l'étude. Deuxièmement, toutes les mesures de précision ainsi que les tests ont été produits en prenant en considération la complexité du plan de sondage de l'étude.

Malgré toutes les précautions prises pour assurer la qualité des données et des tests de même que pour minimiser les biais, il est impossible, comme dans toutes les enquêtes de population, de garantir l'exactitude des réponses fournies par les répondants. Les personnes interrogées peuvent être influencées, entre autres, par le phénomène de désirabilité sociale, par la difficulté de se souvenir des choses passées ou d'évaluer le temps écoulé depuis un événement. De plus, les renseignements portant sur les membres du ménage qui sont recueillis dans le QRI peuvent manquer de précision quand ils proviennent d'une tierce personne.

Une enquête transversale comme *l'Étude auprès des communautés culturelles* permet d'observer des liens entre deux variables ainsi que des différences entre des sous-groupes de la population ou avec des enquêtes similaires passées. Cependant, ce genre d'étude ne permet pas d'établir de lien statistique de causalité entre les caractéristiques étudiées.

De plus, comme la plupart des analyses présentées dans ce rapport s'appuient sur des méthodes bivariées, la prudence est de mise dans l'interprétation de certains résultats pour lesquels le contrôle de certains facteurs exogènes aurait été nécessaire et rendu possible par le recours à la standardisation ou à l'analyse multivariée. L'approche retenue a néanmoins

l'avantage de permettre une bonne description, fort utile en soi, qui constitue par ailleurs une excellente exploration des données recueillies en plus de suggérer des pistes de recherche ultérieure. Quelques analyses multivariées ont toutefois pu être réalisées aux chapitres 8 et 11. Ces analyses ne sont pas exhaustives mais elles apportent de l'information supplémentaire aux analyses bivariées habituelles.

Finalement, la comparabilité des données de l'*Étude auprès des communautés culturelles* avec les données québécoises de l'*Enquête sociale et de santé 1998* a nécessité une attention particulière. Plusieurs questions et indicateurs de cette dernière ont été repris dans la présente étude afin de pouvoir comparer les groupes culturels étudiés avec l'ensemble de la population québécoise dans le but de voir comment ils se situent dans leur nouveau contexte de vie. On a également pris soin de standardiser les données de l'enquête québécoise selon l'âge et le sexe d'après les cinq populations de référence possibles (population de l'ensemble des communautés, populations chinoise, haïtienne, arabophone et hispanophone) pour effectuer adéquatement les comparaisons désirées.

Bibliographie

CLARKSON, M., N. AUDET, R. TREMBLAY, R. COURTEMANCHE et autres (à paraître). *Étude auprès des communautés culturelles 1998-1999, Cahier technique*, Québec, Institut de la statistique du Québec.

BOGUE, D. J., E. E. ARRIAGA et D. L. ANDERTON (1993). *Readings in population research methodology*, Chicago, Illinois, Social Development Center, vol. 1, pages multiples.

COX, B. G. et S. B. COHEN (1985). *Methodological Issues for Health Care Surveys*, New York, Marcel Dekker, 446 p.

DAVELUY, C., L. PICA, N. AUDET, R. COURTEMANCHE, F. LAPOINTE et autres (2000). *Enquête sociale et de santé 1998, 2^e édition*, Québec, Institut de la statistique du Québec, 342 p.

KASS, G. V. (1980). « An Exploratory Technique for Investigating Large Quantities of Categorical Data », *Applied Statistics*, vol. 29, p. 119-127.

KORN, E. L., et B. I. GRAUBARD (1999). *Analysis of Health Surveys*, New York, John Wiley & Sons Inc., 382 p.

LEHTONEN, R., et E. J. PAHKINEN (1996). *Practical Methods for Design and Analysis of Complex Surveys*, Chichester, England, John Wiley & Sons, Ltd, 344 p.

LOHR, S. L. (1999). *Sampling: Design and Analysis*, Pacific Grove, California, Duxbury Press, 494 p.