

Chapitre 2

Méthodologie

Paul Berthiaume

Direction de la méthodologie et des enquêtes spéciales – ISQ

Introduction

L'objectif de la présente enquête est de fournir un portrait fiable et détaillé des comportements des adolescents québécois en matière de tabagisme. Cet objectif s'inscrit dans un contexte plus global de surveillance du tabagisme. Ce dernier aspect a grandement influencé la méthodologie préconisée en ce qui a trait tant au devis qu'aux indicateurs utilisés, afin que l'enquête soit conforme aux standards nationaux et internationaux de surveillance du tabagisme chez les adolescents.

À cet égard, la méthodologie la plus fréquemment utilisée pour documenter les comportements à risque pour la santé chez les adolescents, tels la consommation d'alcool et de drogues, dont le tabac, est l'enquête par questionnaire autoadministré en milieu scolaire. Il s'agit avant tout d'un contexte qui permet de minimiser les biais possibles de sous-déclaration en garantissant l'anonymat des répondants et, par conséquent, de leurs réponses. Ainsi, les jeunes qui expérimentent ou adoptent la cigarette à l'insu de leurs parents peuvent répondre sans craindre que ces derniers ne soient mis au courant de leurs réponses. C'est également une méthodologie qui favorise l'obtention de très bons taux de réponse, essentiels pour la précision des estimations, et qui permet la réalisation de l'enquête à moindre coût. En contrepartie, ce choix nécessite le recours à un plan de sondage complexe entraînant une certaine perte de précision des résultats.

2.1 Procédure d'enquête

L'enquête en milieu scolaire a la particularité d'exiger l'approbation de nombreux paliers décisionnels, soit les commissions scolaires, les directions d'école, les parents – lorsque l'établissement scolaire le requiert – et enfin les élèves. Un consentement a été demandé à chacun de ces groupes d'acteurs en spécifiant le caractère volontaire de

leur participation. Le processus d'obtention des autorisations s'est échelonné sur trois mois.

Dès le mois de mai 2000, des démarches ont été entreprises auprès des commissions scolaires¹ afin d'obtenir la permission de prendre contact avec les écoles sélectionnées. Certaines ont préféré communiquer elles-mêmes avec leurs écoles mais aucune commission scolaire ne s'est opposée à la réalisation de l'étude. Les 159 directions d'école ont ensuite été informées de la nature du projet au moyen d'une lettre d'introduction. Dix jours plus tard, un contact téléphonique était effectué auprès de la direction de chacune des écoles afin de s'assurer de leur participation et de convenir des modalités de réalisation de la collecte des données.

La décision relative à l'approbation parentale relevait de la direction de l'école. Un établissement seulement s'est prévalu de cette prérogative. Dix jours avant la collecte, les parents ont reçu une lettre détaillant la nature du projet et spécifiant le caractère anonyme de l'enquête. Cette lettre était accompagnée d'un formulaire devant être dûment signé par le parent lorsque ce dernier refusait que son enfant participe à l'étude. Le comité d'éthique de la direction Santé Québec avait donné son aval, lors de la première édition de l'enquête en 1998, à l'utilisation d'un consentement passif compte tenu des mesures prises pour garantir l'anonymat des réponses. Seulement cinq parents ont refusé que leur enfant participe à l'étude.

La collecte des données a été menée par le service des activités de collecte de l'Institut de la Statistique du Québec. Elle s'est déroulée entre le 30 octobre et le 1^{er} décembre 2000. L'instrument de collecte consistait en un questionnaire autoadministré distribué aux élèves

1. Pour obtenir la permission de réaliser l'enquête dans les 159 écoles choisies, il a fallu prendre contact avec 53 commissions scolaires.

durant la première demi-heure d'un cours de sciences humaines. Le questionnaire ne comportait aucun code permettant d'identifier l'élève. Un intervieweur se présentait en classe pour distribuer et faire remplir le questionnaire aux élèves. La participation de ceux-ci était libre et volontaire. Les enseignants étaient conviés à demeurer en classe pour maintenir la discipline sans toutefois pouvoir circuler parmi les élèves. Cette mesure visait à garantir la confidentialité des réponses et à minimiser un biais potentiel de sous-déclaration. La plupart des enseignants sont demeurés en classe pendant que les élèves répondaient au questionnaire, ce qui prenait en moyenne 30 minutes. Une fois les questionnaires remplis, l'intervieweur les glissait dans une enveloppe qu'il scellait devant les élèves.

2.2 Plan de sondage

2.2.1 Population visée

La population visée par l'enquête est constituée de tous les élèves de la 1^{re} à la 5^e année du secondaire inclusivement, inscrits dans les écoles québécoises publiques et privées, francophones et anglophones, à l'automne 2000. Sont exclus de la population visée les jeunes qui fréquentent :

- des établissements hors réseau (relevant du gouvernement fédéral ou d'autres ministères provinciaux);
- des écoles autochtones;
- des écoles situées dans des villes de régions éloignées (Parent, Beaucanton, Natashquan, Baie Johan-Beetz, la région Côte-Nord-du-Golfe-Saint-Laurent, l'Île d'Anticosti et les Îles-de-la-Madeleine);
- des écoles composées d'au moins 30 % de jeunes handicapés (les élèves qui fréquentent ces écoles représentent moins de 3 % des élèves de niveau secondaire au Québec);
- les écoles de la région Nord-du-Québec.

À la suite de ces exclusions, l'enquête vise 96 % de la population des élèves qui fréquentent les établissements d'enseignement secondaire québécois. En raison de facteurs logistiques et financiers, les écoles comptant moins de 25 élèves par année d'études ont été exclues de la population échantillonnée, ce qui constitue 1 % des

élèves du secondaire. Ainsi, la population qui a fait l'objet de l'enquête représente 99 % de la population visée. Ces quelque 1 % d'élèves peuvent présenter des comportements tabagiques différents des répondants, mais on présume que leur petit nombre aurait peu d'influence sur les estimations qui seront faites.

2.2.2 Bases de sondage

Deux bases de sondage sont utilisées pour créer l'échantillon. La première, utilisée pour sélectionner de façon aléatoire les écoles pour chaque année d'études, est constituée des fichiers des clientèles scolaires du ministère de l'Éducation du Québec pour l'année scolaire 1999-2000. Les fichiers comprennent entre autres les coordonnées de l'établissement scolaire, le nombre d'étudiants inscrits dans chacune des années du secondaire, le réseau d'enseignement ainsi que la langue d'enseignement.

La seconde base de sondage correspond à la liste des classes pour chacune des années d'études dans les écoles sélectionnées. Les classes appelées à participer sont choisies de façon aléatoire dans cette liste.

2.2.3 Plan d'échantillonnage et stratification

L'échantillon a été construit selon un plan de sondage par grappes² stratifié à deux degrés. Les cinq années du niveau scolaire secondaire constituant des populations indépendantes, la sélection de l'échantillon s'est faite de la façon suivante :

Pour chacune des années d'études :

1. La population des écoles est stratifiée selon la langue d'enseignement (anglais ou français), le réseau d'enseignement (privé ou public) et un découpage géographique en régions métropolitaines de recensement de 1996, lorsque la taille de la population le permettait, tel que le présente la figure 2.1. À l'intérieur de chacune des strates, les écoles ont été sélectionnées de façon aléatoire avec probabilité proportionnelle à leur taille (la probabilité pour une

2. Les personnes échantillonnées sont concentrées dans un endroit précis, en l'occurrence dans une classe, créant ainsi un effet d'agglomération, d'où l'appellation « grappe »

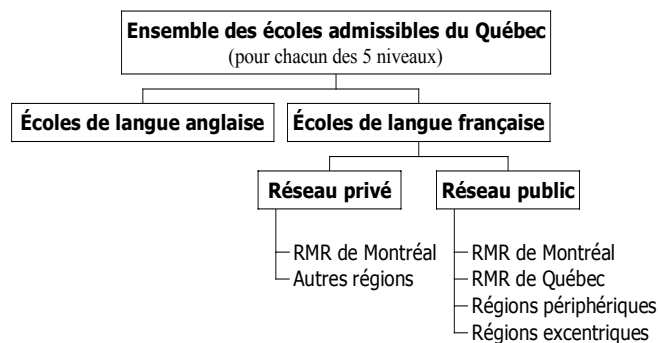
école d'être choisie augmentant avec le nombre d'élèves inscrits dans l'année d'études visée).

2. Ensuite, dans chacune des écoles sélectionnées, une liste des classes de sciences humaines³ a été établie. Une fois les classes d'une école répertoriées, une seule était sélectionnée de façon aléatoire : même probabilité pour toutes les classes d'être choisies. Enfin, tous les élèves de la classe choisie étaient appelés à participer à l'enquête.

On a donc échantillonné une école par année d'études, puis une classe par école. Ainsi, puisque les échantillons par année d'études sont construits de façon indépendante, il est possible qu'une même école ait été sélectionnée dans différentes strates (années d'études).

Figure 2.1

Stratification des écoles pour chaque année d'études



Le plan de sondage de 1998 comprenait un degré d'échantillonnage supplémentaire, correspondant à la sélection de deux régions parmi quatre dites excentriques. Pour l'enquête de 2000, ce premier degré d'échantillonnage a été éliminé dans le but d'améliorer la précision des résultats.

3. Deux raisons justifiaient le choix de cette discipline. Le cours devait être obligatoire pour donner à chaque élève une probabilité non nulle d'être sélectionné. Ensuite, l'échantillon devant être représentatif de l'ensemble des élèves québécois, il fallait éviter les matières soumises à des programmes de performance, telles les mathématiques ou le français. Le bloc des sciences humaines répondait à ces deux critères.

2.3 Taille et répartition de l'échantillon

La taille d'un échantillon est établie en fonction de la précision statistique désirée. Ainsi, pour chacune des années d'études, la taille de l'échantillon devait être suffisamment grande pour produire des estimations d'une précision fiable, c'est-à-dire dont le coefficient de variation (C.V.) est inférieur à 15 % (la notion de coefficient de variation est définie à la section 2.6.2), pour une proportion estimée de 15 % selon le sexe et l'année d'études.

Par ailleurs, le plan de sondage par grappes entraîne une diminution de la précision des estimations obtenues par rapport à un plan de sondage aléatoire simple; en effet, parce que les élèves sélectionnés appartiennent tous à une même classe, ils peuvent présenter une certaine homogénéité au niveau des comportements, d'où la perte de précision encourue. La taille de l'échantillon doit donc être augmentée pour tenir compte de cette perte d'efficacité qui se mesure à l'aide de l'effet de plan. Finalement, un dernier élément à considérer pour déterminer la taille de l'échantillon est le taux de réponse attendu.

Afin d'atteindre les objectifs de précision et considérant les contraintes énumérées ci-dessus, les tailles d'échantillon ont été fixées à environ 1 068 élèves par année d'études (environ 5 340 élèves de niveau secondaire). Cette taille est déterminée selon les hypothèses utilisées pour la planification de l'enquête de 1998, soit un effet de plan inférieur à 1,8 et un taux de réponse combiné, des classes et des élèves, de 85 %. Pour déterminer le nombre d'écoles à sélectionner pour chacune des cinq années d'études, on a utilisé le nombre moyen d'élèves par classe, soit 30 pour les écoles francophones et 27 pour les écoles anglophones⁴. Par conséquent, le nombre d'écoles par année d'études en 2000 est établi à 36, soit 180 pour l'ensemble du secondaire.

4. L'information relative au nombre moyen d'élèves par classe provient des données de l'enquête québécoise sur le tabagisme menée en 1998.

Tableau 2.1

Répartition des tailles d'échantillon des enquêtes de 1998 et de 2000

	1998	2000
Nombre de classes par année d'études	36 (1 ^{re} à 3 ^e sec.) 28 (4 ^e et 5 ^e sec.)	36
Nombre total de classes	164	180
Nombre d'écoles différentes ¹	137	159
Nombre d'élèves possible	4 920	5 340
Nombre d'élèves répondants attendu	3 980	4 540
Proportion minimale estimée pour obtenir un C.V. < 15 %		
• Par année d'études et par sexe	15 % (1 ^{re} à 3 ^e sec.) 20 % (4 ^e et 5 ^e sec.)	15 %
• Par année d'études	12 % (1 ^{re} à 3 ^e sec.) 16 % (4 ^e et 5 ^e sec.)	10 %
• Pour les cinq années d'études et par sexe	6 %	4 %
• Pour les cinq années d'études	3 %	3 %

1. Certaines écoles ont été sélectionnées pour différentes années d'études de façon indépendante.

Source : Institut de la statistique du Québec, *Enquête québécoise sur le tabagisme chez les élèves du secondaire, 2000*

2.4 Modifications entre les enquêtes de 1998 et de 2000

Outre la modification au plan de sondage consistant à éliminer un degré d'échantillonnage (voir section 2.1), un deuxième ajustement a été apporté au nombre de classes sélectionnées en 4^e et 5^e secondaire. En effet, dans l'enquête de 1998, le nombre de classes de 4^e et 5^e secondaire avait été diminué à 28 au profit des classes de 1^{re} à 3^e secondaire qui en comptaient 36 par niveau. Afin d'améliorer la précision des résultats pour l'enquête de 2000, le nombre de classes de 4^e et 5^e secondaire est passé à 36, comme pour chacune des autres années d'études.

Le tableau 2.1 résume les répartitions échantillonnelles des enquêtes de 1998 et de 2000.

2.5 Taux de réponse

Le taux de réponse est défini comme le rapport entre le nombre d'unités répondantes et le nombre d'unités admissibles à l'enquête. Pour la présente enquête, trois taux de réponse ont été calculés : celui des classes, celui des élèves et le taux de réponse combiné des classes et des élèves. Ces taux ont été calculés sur les données pondérées, permettant entre autres la comparaison avec l'enquête de 1998.

2.5.1 Taux de réponse des classes

Des 180 classes sélectionnées initialement, 172 ont accepté de participer à l'enquête, 5 ont refusé de collaborer et 3 n'étaient pas admissibles (école de raccrocheurs, changement de statut de l'école, année d'études n'existant plus). Parce que 3 des 5 classes ayant refusé provenaient du réseau privé, et pour contrer la perte de précision occasionnée par ces refus, il a été décidé de sélectionner 3 classes supplémentaires de façon aléatoire dans ce réseau. Finalement, 175 classes réparties dans 156 écoles ont participé à l'enquête. Ainsi, le taux de réponse des classes s'élève à 97,1 % (voir tableau 2) alors qu'il était en 1998 de 95,4 % (cette dernière valeur a été recalculée afin que les taux de réponse des deux enquêtes soient comparables).

2.5.2 Taux de réponse des élèves

Parmi les écoles visitées, 4 730 élèves ont répondu au questionnaire, ce qui porte leur taux de réponse à 95,2 % (tableau 2). Lors de l'enquête de 1998, 4 238 élèves avaient participé, pour un taux de réponse de 94,1 % (cette dernière valeur a été recalculée afin que les taux de réponse des deux enquêtes soient comparables). Les mêmes raisons qu'en 1998 expliquent la non-réponse globale, soit l'absentéisme et les retardataires.

Tableau 2.2

Nombre de répondants et taux de réponse selon l'année d'études, 1998 et 2000

	1998				2000			
	Classes		Élèves		Classes		Élèves	
	n	taux %	n	taux %	n	taux %	n	Taux %
1 ^{re} secondaire*	34	94,0	944	95,5	36	97,4	1 000	95,2
2 ^e secondaire	36	100,0	984	94,5	34	97,4	949	99,1
3 ^e secondaire*	35	97,9	965	94,8	35	93,8	928	95,6
4 ^e secondaire	27	97,4	763	92,5	35	100,0	948	97,2
5 ^e secondaire*	22	87,7	582	93,0	35	97,1	905	87,8
Total	154	95,4	4 238	94,1	175	97,1	4 730	95,2

* Années d'études pour lesquelles une classe supplémentaire a été échantillonnée en 2000.

Source : Institut de la statistique du Québec, Enquête québécoise sur le tabagisme chez les élèves du secondaire, 2000.

2.5.3 Taux de réponse combiné

Le taux de réponse combiné correspond au produit du taux de réponse des classes et de celui des élèves; il s'établit à 92,4 % pour la présente enquête alors qu'il était de 89,8 % en 1998.

2.6 Traitement et analyse des données

2.6.1 Pondération

La pondération consiste à attribuer à chaque répondant une valeur (un poids) qui correspond au nombre d'individus, incluant lui-même, qu'il représente dans la population visée. Le calcul du poids d'un élève répondant à l'enquête s'effectue selon les étapes suivantes : 1) on détermine la probabilité de sélection de sa classe; 2) on ajuste pour la non-réponse des classes et des élèves; et 3) on ajuste à la population.

De façon plus détaillée, la démarche consiste dans un premier temps à attribuer un poids initial à chaque élève ayant répondu au questionnaire afin de tenir compte de la non-proportionnalité de l'échantillon par rapport à la population visée. Ce poids initial est défini par l'inverse de la probabilité de sélection de sa classe.

Ce poids initial est ensuite ajusté en fonction de la non-réponse des classes et de la non-réponse des élèves. C'est donc dire que le poids initial a été multiplié par l'inverse du taux de réponse des classes par strate ou par

regroupement de strates et par l'inverse du taux de réponse des élèves par classe. Les ajustements pour la non-réponse ont pour conséquence d'augmenter la part de représentativité des élèves répondants, ce qui se traduit par une augmentation des poids initiaux.

Le dernier ajustement apporté au poids correspond à la post-stratification. Cette procédure permet d'ajuster la pondération afin de s'assurer que la distribution pondérée des répondants est conforme à celle de la population visée selon l'année d'études et le sexe. Les données utilisées pour procéder à cet ajustement proviennent du fichier des déclarations scolaires du ministère de l'Éducation du Québec pour l'année 2000-2001.

2.6.2 Estimations

Toutes les estimations fournies dans le présent rapport ont été effectuées en utilisant la pondération et les particularités du plan de sondage. Les estimations de variance ou relatives à la variance (coefficient de variation et intervalles de confiance) ont été calculées à l'aide du logiciel SUDAAN. Pour l'enquête de 1998, des effets de plan moyens avaient été utilisés dans le calcul des estimations relatives à la précision des résultats, selon le type d'estimation et le niveau d'analyse.

L'effet de plan sert à évaluer la perte ou le gain en précision imputable au fait d'avoir eu recours à un plan de sondage complexe (plan de sondage utilisant la notion de degrés d'échantillonnage et/ou de grappes). Il est défini

comme le quotient de la variance estimée avec le plan de sondage complexe par la variance estimée avec un plan de sondage aléatoire simple basé sur le même nombre de personnes.

Pour fins de comparaisons avec l'enquête de 1998, l'effet de plan global pour la présente enquête a été estimé à environ 1,7 alors qu'il était de 1,9 en 1998. Cette amélioration peut s'expliquer par les modifications apportées au plan de sondage (l'élimination d'un degré d'échantillonnage) ainsi que par une meilleure répartition du nombre de classes dans chacune des strates.

2.6.3 Tests statistiques

Les relations entre deux variables, ainsi que les comparaisons de certains résultats avec l'enquête de 1998, ont été étudiées à l'aide de tests du Khi carré (tests d'indépendance et tests d'homogénéité respectivement). Si tous les résultats présentant des différences significatives ne sont pas commentés, toutes les différences mentionnées sont significatives à un seuil de 5 % à moins d'avis contraire. Lorsque les tests d'indépendance ou d'homogénéité, le cas échéant, entre les deux variables à l'étude étaient significatifs, des tests d'égalité entre deux proportions ont été réalisés à l'aide de tests *t* de Student.

2.7 Évaluation méthodologique de l'enquête

2.7.1 Non-réponse partielle

La non-réponse partielle est un phénomène que l'on rencontre dans toute enquête. Elle correspond à la non-réponse, intentionnelle ou non, à une question donnée. Elle peut entraîner des biais dans les estimations si les non-répondants présentent des caractéristiques différentes de celles des répondants. Il est donc important d'analyser la non-réponse partielle à chacune des questions pour évaluer les risques de biais potentiels.

Le taux de non-réponse partielle est défini comme le rapport entre le nombre de personnes ne fournissant pas de réponse à une question et celles qui devaient y répondre. On doit cependant mentionner que les réponses de type « Ne sait pas » peuvent avoir une valeur

informatrice; elles ne sont donc pas traitées en données manquantes et figurent dans les résultats.

De façon générale, on porte une attention particulière aux questions et indices pour lesquels le taux de non-réponse partielle est supérieur à 10 %. Cependant, lorsque le taux de non-réponse partielle est supérieur à 5 % et que le nombre de répondants potentiels est non négligeable, les résultats font également l'objet d'une analyse comparative des caractéristiques des répondants et des non-répondants. Dans la présente enquête, la majorité des résultats présentés ont des taux de non-réponse partielle inférieurs à 5 %. Sinon, ces taux fluctuent entre 5 % et 10 % et les non-répondants ne présentent pas de caractéristiques différentes des répondants.

2.7.2 Erreur d'échantillonnage

Les estimations basées sur un échantillon comportent toujours un certain degré d'erreur lié au fait que l'enquête n'est pas menée auprès de l'ensemble de la population visée. L'ampleur de cette erreur d'échantillonnage est en partie liée au nombre de répondants. La précision des estimations est donc fonction du nombre de répondants à partir duquel elles sont établies. Dans une enquête comme celle-ci, il est important de caractériser la qualité des estimations produites. La marge d'erreur et le coefficient de variation sont deux mesures de précision fréquemment utilisées.

Dans le présent rapport, la mesure principale utilisée pour caractériser la précision des estimations est le coefficient de variation (C.V.). Il permet de mesurer la précision relative des estimations. Il s'exprime comme le rapport, en pourcentage, de l'erreur-type de la proportion estimée sur la proportion estimée elle-même. Pour deux estimations faites sur la même population ou sous-population, la plus petite estimation aura un coefficient de variation plus grand, car plus le phénomène étudié est rare, moins bonne est la qualité de l'estimation produite.

Parce qu'elles sont suffisamment précises, les estimations dont le C.V. est inférieur à 15 % sont présentées dans le rapport sans commentaire; celles dont le C.V. se situe entre 15 % et 25 % sont marquées d'un astérisque (*) et doivent être interprétées avec prudence. Les estimations

dont le C.V. est supérieur à 25 % sont marquées d'un double astérisque (**) pour en signaler l'imprécision et ne sont fournies qu'à titre indicatif.

2.7.3 Portée et limites

La présente enquête est représentative des adolescents québécois inscrits dans un établissement d'enseignement secondaire. Cependant, elle ne peut prétendre à une représentativité de tous les adolescents québécois puisque tous les âges ne sont pas dûment représentés dans la population étudiante de niveau secondaire. La majorité des élèves débutent la 1^{re} secondaire à l'âge de 12 ans, prennent en général un an pour compléter chacune des cinq années et terminent leur secondaire à 16 ou 17 ans selon leur date d'anniversaire. Cependant, le fichier des déclarations scolaires indique qu'en septembre 2000, 19,5 % des élèves de 12 ans sont au primaire, 78 % sont à la 1^{re} secondaire et 2,4 % sont à la 2^e secondaire. La population visée par cette enquête n'est pas la population des jeunes d'un certain groupe d'âge mais plutôt les élèves de niveau scolaire secondaire. Ainsi, les groupes d'âge extrêmes, 12 ans et moins et 17 ans et plus, sont les moins bien représentés.

À l'instar de toutes les enquêtes s'appuyant sur le principe de l'autodéclaration, il est impossible de garantir l'exactitude des réponses fournies par les répondants. Néanmoins, tous les éléments susceptibles d'entraîner un biais de sous-déclaration ont été examinés afin de réduire les risques au minimum. Les précautions suivantes ont été prises pour assurer la validité des données :

- le questionnaire était anonyme;
- le questionnaire était rempli en classe;
- un enquêteur professionnel indépendant de l'école distribuait et récupérait les questionnaires;
- les enseignants ne pouvaient circuler dans la classe pendant que les élèves remplissaient le questionnaire.

Comme on peut le constater, tout a été mis en œuvre pour rassurer les jeunes quant à l'impossibilité pour leurs parents ou l'enseignant de connaître leurs réponses, réduisant ainsi le biais de sous-déclaration. On ne peut cependant exclure la possibilité inverse, soit une

surdéclaration liée à la présence des pairs au moment de remplir le questionnaire.

En terminant, il importe de préciser que le devis transversal de l'enquête permet d'établir des liens ou des associations entre deux variables ainsi que des différences entre des sous-groupes de la population. Toutefois, il ne permet pas l'établissement de relations causales entre les variables.

