

Article

L'évolution de la recherche sur les enquêtes au cours des 25 dernières années

par Graham Kalton

Juin 2000



Statistique
Canada

Statistics
Canada

Canada

L'évolution de la recherche sur les enquêtes au cours des 25 dernières années

Graham Kalton ¹

Résumé

Pour marquer le vingt-cinquième anniversaire de *Techniques d'enquête*, l'auteur passe en revue les principales réalisations de la recherche sur les enquêtes au cours des 25 dernières années. Il présente une synthèse générale de l'évolution de la profession, des méthodes d'enquête (conception des questionnaires, collecte des données, traitement des données manquantes, échantillonnage, erreur d'enquête totale) et des applications (enquêtes par panel, enquêtes internationales, analyse secondaire). Il examine enfin les perspectives d'avenir dans ces secteurs.

Mots clés : Profession de spécialiste des enquêtes; méthodes d'enquête; applications des enquêtes; conception des questionnaires; enquêtes internationales.

1. Introduction

Techniques d'enquête célèbre cette année son vingt-cinquième anniversaire. Pour marquer ce jalon, je passerai en revue, dans le présent article, les principales réalisations de la recherche sur les enquêtes au cours des 25 dernières années. Je tiens à souligner, toutefois, que pour plusieurs raisons les dates risquent d'être un peu floues. Tout d'abord, bien sûr, la recherche sur les enquêtes n'a été marquée d'aucun événement spécial en 1975. Au contraire, bon nombre des progrès majeurs survenus au cours du dernier quart de siècle se sont inspirés de travaux antérieurs. Deuxièmement, dans bien des cas, il faut un certain temps avant que les progrès méthodologiques soient acceptés et adoptés intégralement. Troisièmement, mon point de repère est un texte sur la méthodologie des enquêtes que Sir Claus Moser et moi avons publié au Royaume-Uni en 1971 (deuxième édition de *Survey Methods in Social Investigation*, ci-après intitulée *Survey Methods*), la période couverte s'étendant donc en réalité sur plus de 30 ans.

Le présent exposé passe en revue les progrès de la méthodologie des enquêtes, y compris la conception des questionnaires, l'échantillonnage, les méthodes de collecte des données, le traitement des données et l'analyse des enquêtes. L'informatique est un aspect central, car elle a exercé une influence majeure sur de nombreux progrès méthodologiques, mais pas tous. Le présent exposé examine également l'effet de ces progrès méthodologiques sur les travaux de recherche liés aux enquêtes, y compris l'évolution des enquêtes par panel, des enquêtes internationales et de l'analyse secondaire. J'insisterai surtout sur les enquêtes démographiques, mais je traiterai aussi des enquêtes auprès des établissements. De plus, puisque je me fonde sur mon expérience, l'exposé privilégie sans doute les travaux accomplis aux États-Unis. Avant de me pencher sur les méthodes et la pratique des enquêtes, je compte décrire

d'abord l'énorme accroissement du nombre d'enquêtes et l'établissement explicite de la profession de spécialiste des enquêtes.

2. La profession de spécialiste des enquêtes

L'histoire de la recherche sur les enquêtes relève largement du XX^e siècle. Cette activité a pris son essor dans les années 1930, a connu une croissance considérable au cours de la Seconde Guerre mondiale et a maintenu ensuite un taux de croissance appréciable. En 1975, les enquêtes auprès des ménages et des établissements étaient déjà bien établies comme moyen de répondre aux besoins en données statistiques des décideurs et des chercheurs de nombreux secteurs comme le commerce et la fabrication, l'agriculture, l'emploi et le chômage, les dépenses des familles, la nutrition, la santé, l'éducation, les voyages, le vieillissement et la criminalité. De plus, les enquêtes menées par les universitaires et d'autres chercheurs en sociologie, en science économique, en science politique, en psychologie, en éducation, en travail social et en hygiène publique, en sondages d'opinion et en sondages électoraux de même qu'en études de marché se sont multipliées. Ce secteur a continué de se développer rapidement au cours des 25 dernières années, surtout que les décideurs ont appris à reconnaître la valeur des données d'enquête, tandis que le perfectionnement des méthodes d'enquête a permis aux spécialistes de répondre à la demande de données statistiques. La demande soutenue de données de plus en plus détaillées parmi les décideurs a stimulé le perfectionnement des méthodes d'enquête tout en favorisant l'établissement d'une profession bien implantée de spécialistes des enquêtes.

La croissance rapide de la recherche sur les enquêtes est attribuable en partie à un élargissement du choix de thèmes

1. Graham Kalton, Westat, 1650 Research Boulevard, Rockville, Maryland, 20850, États-Unis. Courriel : KaltonG1@westat.com.

jugés propices. Des chercheurs aventureux ont réussi à repousser les limites conventionnelles quant aux sujets abordés dans les enquêtes. Cette tendance s'est poursuivie au cours des 25 dernières années, de sorte qu'aujourd'hui très peu de sujets sont exclus d'enquêtes fondées sur des échantillons probabilistes valides. Certains nouveaux sujets d'étude sont délicats, par exemple le comportement sexuel et l'utilisation de drogues illicites, et l'application des méthodes d'enquête a exigé l'élaboration de techniques spéciales de collecte des données. D'autres sujets nouveaux ont supposé l'incorporation de méthodes de collecte supplémentaires, par exemple des examens médicaux pour les unités échantillonnées, l'enregistrement vidéo des interactions entre enseignants et étudiants en salle de classe et la mise en place d'équipement de surveillance de l'environnement dans des ménages échantillonnés. Le défi posé par les thèmes difficiles a constamment stimulé la recherche méthodologique.

Avant 1975, il n'existait aucune revue à grand tirage spécialisée en méthodes d'enquête. Les articles évalués traitant de méthodes d'enquête étaient publiés dans différentes revues. Les revues de statistique publiaient, et continuent de publier, des articles traitant surtout des aspects statistiques de la recherche sur les enquêtes, notamment l'échantillonnage. Des revues comme *Public Opinion Quarterly* publiaient, et continuent de publier, des articles sur les méthodes d'enquête. Les revues d'études de marché publient des articles sur les méthodes d'enquête se rapportant aux études de marché. Les revues traitant de disciplines comme les sciences sociales et l'hygiène publique publient parfois des articles sur les méthodes d'enquête se rapportant à leur discipline. Cette situation n'était pas idéale, puisqu'il n'y avait aucun débouché naturel pour de bons articles traitant des méthodes de recherche sur les enquêtes, et parce que la documentation était éparpillée. Le lancement de *Techniques d'enquête* en 1975 et du *Journal of Official Statistics* en 1985 a permis de rectifier la situation; ces deux revues sont maintenant bien établies.

Un autre phénomène important a été l'établissement d'associations professionnelles pour les méthodologistes d'enquête. Ainsi, l'Association internationale des statisticiens d'enquêtes (AISE) a été fondée en 1975 à titre de section de l'Institut international de statistique. La Section on Survey Research Methods de l'American Statistical Association a été établie en 1978, après avoir été une sous-section de la Social Statistics Section entre 1974 et 1977. La Social Statistics Section de la Royal Statistical Society a été fondée en 1976, d'abord sous le nom de Social Statistics and Survey Methodology Study Group.

Depuis quelques années, plusieurs de ces associations, œuvrant parfois avec d'autres associations, notamment l'American Association for Public Opinion Research, ont participé à l'organisation de conférences internationales sur différents aspects des méthodes d'enquête. À noter que plusieurs de ces conférences ont été structurées de façon à

traiter un thème de façon globale et de produire des textes bien équilibrés. Il s'agissait de combler les lacunes documentaires suscitées par le fait que les méthodologistes d'enquête sont des praticiens qui n'ont guère le temps de publier. Il en est résulté la préparation de recueils publiés sur des thèmes comme les enquêtes par panel, les enquêtes téléphoniques, les enquêtes auprès des entreprises, les erreurs de mesure liées aux enquêtes, la qualité des enquêtes et la collecte assistée par ordinateur.

De nombreuses autres conférences sur les méthodes d'enquête ont eu lieu. Les unes ont été organisées par des organismes publics comme Statistique Canada, le Bureau of the Census des États-Unis et le Federal Committee on Statistical Methodology des États-Unis, fondé lui aussi en 1975. Les autres ont été organisées par des associations professionnelles comme l'AISE et l'Association for Survey Computing. Les actes de ces conférences et ceux de la Section on Survey Research Methods de l'American Statistical Association ont largement favorisé l'accroissement de la documentation sur les méthodes d'enquête.

Deux autres aspects de l'évolution de la profession de spécialiste des enquêtes méritent d'être soulignés. Le premier est son caractère international. Les conférences internationales décrites ci-dessus ont donné lieu à des publications d'auteurs de plusieurs pays. Malgré l'existence de différences culturelles entre pays, dont il faut tenir compte dans la collecte des données, la recherche sur les méthodes d'enquête comporte de nombreux aspects communs à plusieurs pays. De plus, les enquêtes internationales sont plus fréquentes, d'où le besoin de normaliser les procédures d'un pays à l'autre (voir ci-dessous). De façon générale, la coopération internationale en matière de recherche sur les enquêtes va bon train, mais il existe un secteur qui mérite une bien plus grande attention. Tout comme les pays développés, les pays en développement et en transition ont besoin de données statistiques tirées d'enquêtes. Toutefois, ils n'ont pas toujours les compétences nécessaires. L'AISE, des organismes internationaux comme le Bureau de statistique des Nations Unies, plusieurs bureaux de la statistique du secteur public et d'autres organismes jouent un rôle important dans la formation de spécialistes des enquêtes dans les pays en développement et en transition, mais le soutien prévu actuellement à cet égard est loin de combler les besoins.

L'autre aspect de l'évolution de la profession de spécialiste des enquêtes est son caractère multidisciplinaire. En trouvant sa place parmi les disciplines professionnelles, la recherche sur les enquêtes a donné lieu à des sous-disciplines. Il y a une trentaine d'années, un méthodologiste des enquêtes pouvait s'attendre à couvrir tous les aspects de sa matière, mais cela n'est plus possible au niveau technique le plus élevé. La technicité de l'échantillonnage et de l'analyse des enquêtes est très poussée, et les méthodologistes des enquêtes utilisent de plus en plus des théories et des techniques se rapportant à la sociologie, à la psychologie et à l'anthropologie, tandis que les informaticiens

doivent avoir recours à des méthodes beaucoup plus perfectionnées que par le passé pour la saisie et le traitement des données. Cette segmentation inévitable des méthodes d'enquête, compte tenu des progrès accomplis, risque de nuire à l'unité de la profession, d'autant plus que les sous-disciplines donnent naissance à leur tour à des sous-secteurs. Vu l'importance d'une collaboration interdisciplinaire pour ce qui est de la recherche sur les enquêtes, il va peut-être falloir trouver à l'avenir des mécanismes favorisant une telle collaboration (voir la section 5).

Tout comme les pays en développement et en transition, les pays développés n'ont pas suffisamment de méthodologistes et de statisticiens d'enquête bien formés. Il importe à la fois d'attirer plus de gens à la profession et de prévoir plus de possibilités de formation. Il existe quelques programmes d'études supérieures dans les universités, et certains professeurs se spécialisent dans ce secteur, mais les effectifs sont insuffisants compte tenu des besoins. La collaboration multidisciplinaire nécessaire à la préparation et à l'exécution d'une enquête suppose que la formation comporte un volet multidisciplinaire, afin que les spécialistes puissent communiquer de façon efficace entre eux. De plus, parmi les instructeurs on devrait trouver des personnes ayant une expérience pratique des enquêtes. Compte tenu de ces exigences, il est difficile de mettre sur pied un programme d'études supérieures en méthodes d'enquête dans la plupart des universités. Une autre stratégie a été adoptée dans le cadre du Joint Program in Survey Methodology (JPSM) de l'université du Maryland, mis sur pied grâce à des fonds publics aux États-Unis pour combler le manque de spécialistes d'enquête dûment formés dans l'administration fédérale. Ce programme se fonde sur une collaboration entre deux universités (l'université du Maryland et l'université du Michigan) et un organisme privé de recherche sur les enquêtes (Westat), avec la participation notable de spécialistes en méthodes d'enquête du secteur public, d'autres organisations et d'autres universités. De même, le Department of Social Statistics de l'université de Southampton et l'Office for National Statistics du Royaume-Uni ont récemment élaboré un programme de deuxième cycle universitaire en statistique officielle, avec la participation, pour l'enseignement des méthodes d'enquête et d'autres aspects de la statistique officielle, de statisticiens du secteur public. Ce département d'université collabore également avec un organisme indépendant de recherche sur les enquêtes (le National Centre for Social Research) dans le cadre du Centre for Applied Social Surveys, dont une des activités est d'offrir des cours abrégés en méthodes d'enquête.

3. Évolution des méthodes d'enquête

La révolution informatique qui a commencé à influencer les analyses d'enquête dans les années 1960 a joué un rôle dominant dans l'évolution des méthodes d'enquête au cours

des 25 à 30 dernières années. La capacité de traiter et d'analyser les données d'enquête beaucoup plus aisément que par le passé a favorisé l'utilisation de méthodes statistiques plus avancées. Elle a également suscité des demandes beaucoup plus détaillées de la part des utilisateurs des données d'enquête, stimulant la mise au point de méthodes améliorées pour tous les aspects du processus d'enquête.

Le chapitre sur le traitement des données d'enquête dans *Survey Methods* contient une description des cartes perforées qui étaient utilisées largement il y a trente ans pour l'analyse des données d'enquête, de même que du matériel d'enregistrement (compteuses-trieuses et tabulateurs) et des ordinateurs. À cette époque, les ordinateurs étaient à la veille de remplacer le matériel d'enregistrement, mais ils n'étaient pas habituellement accessibles aux spécialistes des enquêtes. Les ordinateurs de l'époque étaient de gros appareils centraux, et les cartes perforées étaient l'outil habituel de saisie des données d'enquête. Le nombre et l'envergure des programmes d'analyse des enquêtes étaient limités. De nos jours, la situation est tout à fait différente, et on ne saurait surestimer l'importance de ce changement pour la recherche sur les enquêtes.

C'est dans le contexte de cette explosion informatique qu'il convient d'évaluer les progrès réalisés dans d'autres secteurs de la méthodologie des enquêtes. Le reste de la présente section décrit brièvement les progrès importants qui, à mon avis, ont été réalisés au cours du dernier quart de siècle pour ce qui est de la conception des questionnaires, de la collecte des données, des données manquantes, de l'échantillonnage et de l'erreur d'enquête totale.

Conception des questionnaires. Le rôle essentiel de la conception des questionnaires dans l'obtention de données d'enquête de qualité élevée a été reconnu dès le début. Il est vrai que d'excellents travaux ont été menés sur l'amélioration de la conception des questionnaires au cours des années 1960 et 1970, mais le nombre de chercheurs qui se penchaient sur cette tâche très épineuse était très limité. La situation s'est améliorée appréciablement à cause surtout de ce que l'on a appelé le mouvement CASM (aspects cognitifs des méthodes d'enquête). Le mouvement CASM cherche à recruter des spécialistes des sciences cognitives et sociales pour aborder les problèmes épineux de la formulation des questions d'enquête pouvant susciter des réponses appropriées. L'intérêt suscité par ce mouvement a entraîné un nouvel essor dans ce secteur.

Le mouvement CASM n'a pas mis en évidence des solutions toutes faites aux problèmes des erreurs de réponse dans les enquêtes. Il n'aurait pas été raisonnable de s'attendre à tout régler par l'importation de théories existantes de la psychologie cognitive et d'autres disciplines. Le mouvement a permis de multiplier les efforts dans une perspective théorique. De plus, le mouvement CASM a permis d'établir des essais préliminaires plus rigoureux pour les questionnaires d'enquête. Certaines techniques d'essai préliminaire élaborées au cours des 25 dernières années sont survenues indépendamment du mouvement CASM, mais

l'intérêt soutenu que l'on manifeste pour les essais préliminaires est attribuable en grande partie à ce mouvement. Un résultat immédiat du mouvement CASM a été la création des « laboratoires cognitifs » dont on se sert beaucoup maintenant pour les essais préliminaires de questionnaires à l'aide de techniques comme la réflexion à haute voix et les questions exploratoires. Les groupes de discussion, associés depuis longtemps à la conception des questionnaires, surtout pour les études de marché, sont eux aussi utilisés beaucoup plus souvent que par le passé. De plus, le codage du comportement est utilisé largement à l'heure actuelle pour les essais préliminaires.

Un phénomène récent connexe a été l'adoption d'une stratégie plus théorique pour la conception des formulaires d'enquête que les répondants doivent remplir. Ce genre de recherche tient compte de théories qui expliquent comment les particuliers abordent un document et comment ils le consultent. Ce domaine important a suscité peu d'intérêt pendant plusieurs années. Les recherches actuelles sont très prometteuses et devraient rendre les formulaires d'enquête plus conviviaux, ce qui devrait favoriser une amélioration de la qualité des données recueillies et des taux de réponse.

Collecte des données. *Survey Methods* contient deux chapitres qui traitent des méthodes de collecte des données : le premier aborde les questionnaires postaux et le deuxième les interviews directes (un chapitre décrit aussi les documents et l'observation). On ne mentionne que brièvement les interviews téléphoniques, en partie à cause de la faible pénétration du téléphone au Royaume-Uni à cette époque. Toutefois, même aux États-Unis où la pénétration du téléphone était beaucoup plus marquée, en 1975 de nombreux spécialistes des enquêtes avaient de sérieuses réserves au sujet de la collecte téléphonique des données pour les enquêtes-ménages, du moins pour les enquêtes gouvernementales comportant d'importantes répercussions en matière de politique. La situation a beaucoup évolué. De nos jours, de nombreuses enquêtes du gouvernement des États-Unis se font par téléphone.

Une préoccupation suscitée par les enquêtes téléphoniques est la non-couverture des ménages sans téléphone. Puisque la couverture téléphonique se rapproche actuellement de 95% aux États-Unis, la non-couverture des ménages sans téléphone peut être considérée comme acceptable pour les enquêtes auprès de la population générale. Toutefois, un nombre important d'enquêtes se penchent sur des sous-populations dont le taux de couverture téléphonique est moins élevé, par exemple les personnes pauvres; la non-couverture téléphonique est une préoccupation sérieuse pour de tels ménages. La non-réponse est elle aussi inquiétante. Les taux de non-réponse pour les enquêtes téléphoniques sont appréciablement plus élevés que pour des enquêtes comparables fondées sur des interviews directes, et l'écart semble s'élargir. Lorsqu'il faut faire un choix entre des modes de collecte téléphoniques et directs, les économies considérables réalisées à l'aide des interviews téléphoniques l'emportent souvent sur le taux de réponse

plus élevé que l'on peut atteindre grâce à des interviews directes. Or, le risque de biais appréciable associé aux niveaux élevés de non-réponse des enquêtes téléphoniques (pouvant souvent atteindre 40% ou plus, même en présence d'un suivi dynamique) représente un problème sérieux et souvent sous-estimé. Puisqu'il est probable que le taux de non-réponse des enquêtes téléphoniques va augmenter, on peut se poser des questions au sujet du rôle de la collecte téléphonique à l'avenir.

Un progrès important a été réalisé pour ce qui est de la collecte des données grâce à la récente introduction de méthodes assistées par ordinateur, par exemple les interviews sur place assistées par ordinateur (IPAO) et les interviews téléphoniques assistées par ordinateur (ITAO). Ces méthodes prévoient des instructions de type « passez à » plus complexes, empêchent les intervieweurs de dévier de la séquence prévue du questionnaire, facilitent les insertions à partir de questions antérieures (exemple : si « Pierre » a été inscrit comme nom du fils lors d'une question antérieure, « Pierre » peut faire partie du libellé d'une question subséquente) et rendent possibles des vérifications à mesure que l'interview se poursuit, des corrections étant apportées au besoin. Le fait d'entrer les données directement dans un fichier informatique assure également un traitement plus rapide. L'élaboration de programmes généralisés pour la collecte de type IPAO et ITAO, y compris l'échantillonnage et la planification des interviews, représente un travail complexe. Plusieurs programmes peuvent être utilisés à cette fin. L'avenir nous réserve sans doute des programmes plus souples et des systèmes auteurs plus simples d'application.

Depuis quelques années, un autre type de collecte des données d'enquête assistée par ordinateur a vu le jour. Il s'agit de l'auto-interview assistée par ordinateur (AIAO), dont il existe plusieurs variantes : l'AIAO-vidéo (le répondant lit les questions à l'écran et inscrit les réponses au clavier), l'AIAO-audio (le répondant entend les questions grâce à des écouteurs reliés à un ordinateur portatif et inscrit les réponses au clavier), l'AIAO-audio téléphonique (l'interview AIAO-audio se déroule par téléphone, et le répondant appelle l'ordinateur ou est transféré à l'interview par ordinateur lorsque la communication téléphonique a été établie par l'intervieweur). Toutes ces versions de l'AIAO évitent l'interaction répondant-intervieweur des autres méthodes d'interview, et pourraient donc être d'une utilité particulière pour la collecte de données sur des sujets délicats. Au besoin, on pourrait également prévoir des versions autres qu'anglaise. Les variantes audio n'exigent pas que le répondant sache lire. Ces méthodes ne sont apparues que récemment, et on peut en prévoir l'essor à l'avenir.

Il existe actuellement des enquêtes-entreprises menées à l'aide de méthodes AIAO-audio. Un avantage est que le répondant peut composer un numéro de libre-appel à un moment qui lui convient. Il entend alors des questions d'enquête transmises par numérisation de la parole, et

inscrit la réponse à l'aide du clavier de son téléphone. Dans une variante de cette méthode, les réponses orales sont interprétées à l'aide de techniques de reconnaissance de la parole. Le recours à cette méthode pourrait s'accroître à mesure que l'on perfectionne les techniques de reconnaissance de la parole.

Un autre progrès récent a été la collecte de données d'enquête par Internet. Une telle façon de procéder est particulièrement intéressante pour certains types d'enquêtes auprès des établissements et pour des enquêtes auprès de personnes ayant accès à Internet et sachant s'en servir. On peut par exemple transmettre le questionnaire par courriel, si l'on connaît l'adresse électronique (comme c'est le cas des employés d'une entreprise ayant son propre réseau). On peut également monter le questionnaire sur un site Web, quitte à ce que le répondant utilise un mot de passe pour y avoir accès. Pour le moment, Internet ne se prête pas aux enquêtes menées auprès du grand public, à cause de la forte proportion de personnes n'y ayant pas facilement accès, de l'absence de plan d'échantillonnage et de taux de réponse probablement faibles. On ne devrait pas succomber à la tentation de recueillir un grand échantillon de réponses Internet à un questionnaire d'enquête en l'absence d'un contrôle approprié. Une telle démarche ne ferait que répéter les erreurs du fameux Literary Digest Poll de 1936.

Données manquantes. Les enquêtes comportent des données manquantes à cause de la non-réponse totale, de la non-réponse partielle et de la non-couverture. Au cours des 25 dernières années, et même auparavant, l'accroissement du taux de non-réponse totale a suscité de plus en plus de préoccupations. Il est difficile de documenter cette tendance, et l'analyse des données de différentes enquêtes ont même suscité des conclusions divergentes quant à l'existence d'une tendance. Et pourtant les responsables des enquêtes sont généralement d'accord qu'il est devenu plus difficile d'obtenir la coopération des gens. Plusieurs explications ont été proposées à cet égard, par exemple le manque de nouveauté des enquêtes, le nombre accru de gens ayant moins de temps libre, la crainte de violence dans les interviews directes, de même que les effets négatifs du télémarketing sur les enquêtes téléphoniques, mais il n'y a pas d'explication définitive. Peu importe les raisons, il faut désormais déployer plus d'efforts que par le passé pour obtenir un taux de réponse élevé. Ainsi, il faut augmenter le nombre d'appels pour communiquer avec les répondants, consacrer plus d'énergie à convertir les gens qui refusent, et faire davantage appel à l'incitation. Au cours de la dernière décennie, on a mené un nombre appréciable d'études expérimentales, lors d'enquêtes par interview directe et téléphonique, afin de vérifier l'effet, sur le taux de réponse, de diverses incitations monétaires et autres; ces études reprenaient dans le contexte d'une interview les analyses menées antérieurement sur les questionnaires postaux.

La non-couverture est un aspect inquiétant des enquêtes téléphoniques, mais elle a suscité moins d'intérêt pour ce qui est des enquêtes par interview directe, et certainement

moins d'attention que le problème de la non-réponse. Et pourtant la non-couverture liée aux enquêtes par interview directe parmi certains segments de la population (les jeunes hommes noirs aux États-Unis par exemple) est parfois élevée. De plus, les unités non couvertes sont peu connues, si ce n'est que l'on peut s'attendre à des différences appréciables relativement aux unités couvertes. C'est là une source d'erreur d'enquête sur laquelle il y aurait lieu de se pencher à l'avenir. La non-couverture peut être particulièrement grave dans une enquête auprès d'une population rare (les adolescents par exemple) lorsque les unités sont échantillonnées dans le cadre d'un processus de sélection de grande envergure. Puisque les enquêtes menées auprès de populations rares suscitent un intérêt accru, ce type de non-couverture mérite une attention particulière.

Il y a vingt-cinq ans, pour la non-réponse partielle, on laissait simplement tomber les cas de l'analyse en question, en calculant par exemple le pourcentage de répartition pour le sous-ensemble de cas comportant des réponses acceptables. On supposait implicitement que le volet non-réponse partielle manquait complètement au hasard (MCAH). On applique toujours cette façon de procéder à de nombreuses enquêtes, mais de plus en plus on a recours à une certaine forme d'imputation pour attribuer des valeurs aux réponses manquantes de façon à tenir compte des réponses fournies aux autres questions d'enquête. Cette stratégie remplace une hypothèse MCAH parfois intenable par une hypothèse de non-réponse manquant au hasard (MAH), voulant que les non-réponses partielles manquent au hasard suivant les variables auxiliaires utilisées dans l'imputation. Il est vrai que l'on utilisait il y a vingt-cinq ans des méthodes d'imputation à l'occasion, mais la plupart des documents importants parus sur ce thème sont postérieurs à 1975. Les méthodes courantes s'appuient fortement sur la puissance actuelle des ordinateurs. L'imputation demeure un domaine de recherche actif comportant deux points de convergence : l'élaboration de méthodes d'imputation conservant la structure de covariance de l'ensemble des données d'enquête, sans oublier que presque toutes les variables d'enquête sont exposées à une non-réponse partielle, et le calcul d'estimations de la variance pour des estimations d'enquête fondées sur des données dont certaines sont imputées (voir ci-dessous).

La vérification des données, qui est liée étroitement à l'imputation, a également connu des progrès appréciables ces dernières années, la puissance accrue des ordinateurs ayant permis de mettre au point des procédures plus complexes que par le passé. Comme l'imputation, la vérification fait l'objet de beaucoup de recherches, et l'avenir est prometteur.

La puissance accrue des ordinateurs est également un facteur important du développement et de la fréquence d'utilisation des corrections pondérées de la non-réponse et de la non-couverture. Les corrections par classe de pondération pour la non-réponse et la non-couverture (stratification a posteriori) s'appliquaient lorsque l'on utilisait le matériel d'enregistrement pour l'analyse des enquêtes, mais

les méthodes étaient forcément assez simples. À l'heure actuelle, on utilise largement des méthodes plus complexes de classe de pondération et de calage, englobant de nombreuses variables auxiliaires, souvent après avoir utilisé des analyses exploratoires pour cerner des variables auxiliaires appropriées.

Échantillonnage. Les principales méthodes liées au plan d'échantillonnage (stratification, échantillonnage à plusieurs degrés, échantillonnage à probabilités inégales, par exemple) ont été élaborées au cours des premières années et décrites dans des manuels publiés dans les années 1950. Les réalisations du dernier quart de siècle ont permis de raffiner et d'élargir ces méthodes, un exemple étant l'échantillonnage à composition aléatoire pour les enquêtes téléphoniques. Encore une fois, c'est la capacité de l'ordinateur de traiter d'énormes volumes de données de recensement et autres grandes bases d'échantillonnage qui a permis aux spécialistes des enquêtes de préparer des plans plus efficaces que par le passé.

Un domaine de recherche, ces dernières années, a été l'étude des méthodes d'échantillonnage pour les populations rares, à l'aide d'une enquête spéciale ou d'une enquête générale avec suréchantillonnage. Il s'agit là d'un volet de l'élargissement de la demande en fonction de résultats pour de nombreux domaines, y compris les plus petits comme les minorités raciales et ethniques, les enfants pauvres, les groupes d'âge et de sexe et les subdivisions géographiques (voir ci-dessous la mention des estimations régionales). Les chercheurs se penchent sur des plans d'échantillonnage et des méthodes de collecte efficaces pour l'échantillonnage de domaines de ce genre dans des situations où il n'existe pas de bases de sondage spéciales. Compte tenu de la croissance ininterrompue de la demande de résultats par domaine, il faut continuer de trouver des façons de sonder des populations rares de façon rentable.

Dans les années 1970, le mode d'inférence fondé sur le plan adopté généralement pour les sondages a été fortement remis en question par des personnes voulant qu'il soit remplacé par les méthodes fondées sur un modèle que l'on utilise ailleurs en statistique. Le débat s'est atténué, et la stratégie fondée sur le plan de sondage est encore utilisée (voir ci-dessous). À cet égard, il y a lieu de clarifier la terminologie : dès le début, le mode d'inférence fondé sur le plan de sondage faisait appel à des modèles afin d'accroître la précision des estimations (par exemple les estimations de régression), mais les estimations demeuraient cohérentes pour ce mode d'inférence indépendamment de la validité du modèle. On distingue donc les procédures assistées par modèle des procédures dépendant d'un modèle. La pertinence des estimations dépendant d'un modèle est fonction de la validité du modèle (ou de la robustesse des estimations vis-à-vis des lacunes du modèle). Les récents progrès de l'informatique ont favorisé un recours accru à des modèles, et à des modèles plus complexes, dans le cadre du mode d'inférence fondé sur le plan de sondage et assisté par modèle.

Ces remarques n'excluent aucunement les méthodes dépendant d'un modèle de la recherche sur les enquêtes. Au contraire, les méthodes de traitement des données manquantes décrites ci-dessus dépendent nécessairement d'un modèle. De plus en plus, des méthodes dépendant d'un modèle servent également à préparer des estimations pour de petits domaines (de petites régions géographiques en général). On a besoin de méthodes de ce genre lorsque la taille des échantillons d'un domaine est trop petite (elle est parfois nulle) pour que l'on puisse préparer des estimations fondées sur le plan de sondage qui soient suffisamment précises. Dans une telle situation, on peut produire des estimations régionales en bénéficiant de données d'enquête d'autres secteurs ou d'autres périodes de temps grâce à un modèle statistique qui relie les données d'enquête à d'autres données, généralement administratives. La croissance rapide de programmes sociaux assurant la répartition de fonds parmi de petites entités géographiques a entraîné une demande considérable d'estimations régionales à jour. Par conséquent, l'estimation régionale est devenue un important secteur de recherche depuis quelques années, et le demeurera probablement.

L'estimation de la variance pour des plans d'échantillonnage complexes a été un autre secteur important depuis un quart de siècle. Des méthodes fondées sur les approximations de série de Taylor et les méthodes de répétition ont été utilisées dans les années 1960, mais leur application n'était pas chose courante et se limitait surtout aux activités de recherche. Cette situation a changé de façon marquée sous l'effet de la puissance accrue des ordinateurs et de la mise au point de logiciels de calcul des erreurs d'échantillonnage pour des estimations de plan d'échantillonnage complexe (typiquement stratifié à plusieurs degrés). De nos jours, le calcul des erreurs d'échantillonnage est une pratique assez courante dans l'analyse des données d'enquête.

Un phénomène récent important a été l'application de modèles analytiques aux données d'enquête. Le débat se poursuit quant au choix entre un mode d'inférence fondé sur le plan de sondage et un mode d'inférence dépendant d'un modèle. Pour ce qui est de la stratégie fondée sur le plan de sondage, on a pu observer des progrès théoriques à la fois dans l'application de modèles de régression, de modèles catégoriques, de modèles de survie, de modèles à plusieurs niveaux, etc. avec des données d'enquête et dans l'utilisation de logiciels pour le calcul de la variance de ces modèles. Présentement, les analystes d'enquête mènent souvent leurs études exploratoires à l'aide de modules statistiques standard plus souples et ils calculent la variance en fonction du plan de sondage en se servant, durant les étapes finales seulement de leurs analyses, d'un logiciel d'estimation de la variance de l'échantillonnage. À l'avenir, les procédures d'estimation de la variance de l'échantillonnage devraient être davantage intégrées à des logiciels standard.

Un secteur très actif actuellement est le calcul de la variance d'estimations d'enquête fondées sur des réponses

dont certaines sont imputées. Une stratégie consiste à appliquer des procédures d'imputation multiples à des plans d'échantillonnage complexes, démarche qui fait largement appel à de puissants ordinateurs. On élabore d'autres méthodes dans le cadre du mode d'inférence fondé sur le plan de sondage standard (forcément en fonction d'hypothèses de modèle). Ces méthodes seront peut-être incorporées un jour dans les programmes d'estimation de la variance de l'échantillonnage en vue d'une application aisée.

Erreur d'enquête totale. La discussion a porté jusqu'à présent sur les diverses composantes du processus d'enquête. Toutefois, une enquête bien conçue est la réunion de composantes en un module efficace tenant compte du facteur coût. On a mieux reconnu cet aspect au cours des 25 dernières années, une attention accrue étant accordée au concept de l'erreur d'enquête totale et à celui du plan d'enquête total. Compte tenu de la réduction des ressources, un plan d'enquête représente un compromis entre, par exemple, la taille de l'échantillon, l'ampleur de la conversion de la non-réponse, la longueur du questionnaire et la qualité des données obtenues à l'aide de différents modes de collecte des données. Au moment d'analyser les données d'enquête, il convient d'évaluer la qualité des estimations en fonction de l'erreur d'enquête totale pour toutes les sources, et non pas simplement l'erreur d'échantillonnage. Pour le plan aussi bien que pour l'analyse, il faut des renseignements détaillés sur les différentes sources d'erreur et leur effet sur les estimations d'enquête. De plus, puisque les enquêtes sont des études complexes comportant de nombreux objectifs analytiques, les besoins en données sont considérables. L'abondante documentation sur les erreurs d'enquête provenant de différentes sources facilite l'étude de l'erreur d'enquête totale et du plan d'enquête total en fonction des contraintes budgétaires, mais il y a lieu de poursuivre la recherche.

Les notions d'erreur d'enquête totale et de plan d'enquête total s'appliquent surtout à des enquêtes répétées. L'information sur les sources d'erreur se laisse accumuler d'un cycle à l'autre, de sorte que l'on peut alors l'utiliser pour établir les priorités quant à l'amélioration des méthodes d'enquête. Les profils qui donnent un aperçu intégré des sources d'erreur dans une enquête (voir ci-dessous) servent notamment à orienter le choix des priorités pour l'amélioration des méthodes.

4. Autres réalisations

On trouvera ci-dessous un aperçu de divers domaines de la recherche sur les enquêtes, autres que ceux de nature méthodologique dont il a été question à la section 3, qui ont donné lieu à d'importants progrès au cours des 25 dernières années. Cet aperçu n'est pas exhaustif; on y traite seulement des secteurs qui, selon moi, ont subi un changement majeur.

Enquêtes par panel. Les avantages des données longitudinales obtenues dans le cadre d'enquêtes par panel sont

reconnus depuis longtemps, et des enquêtes par panel étaient menées au cours des années 1940 et 1950. À cette époque, toutefois, la création d'ensembles de données longitudinales, intégrant des données de divers cycles, était fort complexe. Le fait que l'analyse des enquêtes par panel se faisait souvent largement de façon transversale était une source majeure de critiques de la méthode. De nos jours, les progrès de l'informatique et des techniques d'analyse longitudinale ont modifié la situation considérablement. Néanmoins, la complexité des données longitudinales et, surtout, le problème des données manquantes n'ont pas disparu. À l'heure actuelle, on a largement recours à des méthodes longitudinales d'analyse, même si l'analyse de nombreuses enquêtes par panel se fait toujours largement de façon transversale, trop peu d'attention étant accordée à l'éventail de facettes que leurs données longitudinales pourraient élucider.

La croissance des enquêtes par panel au cours des 20 dernières années a été énorme, englobant tout un choix de sujets, y compris l'éducation, les transitions des travailleurs, la santé et le comportement aux élections. Les enquêtes par panel sur l'économie des ménages, qui s'inspirent de la Panel Study of Income Dynamics lancée en 1968 par l'université du Michigan, ont gagné en popularité et se poursuivent dans plusieurs pays. Mentionnons également l'Enquête sur la dynamique du travail et du revenu de Statistique Canada et la Survey of Income and Program Participation du Bureau of the Census des États-Unis, qui font appel à des méthodes semblables.

Il est tout probable que le recours à des plans de sondage par panel sera encore plus fréquent à l'avenir. Le défi consiste à tirer pleinement profit des données longitudinales résultantes, car le potentiel analytique d'une enquête par panel augmente de façon exponentielle avec le nombre de cycles de collecte de données. De plus, les techniques perfectionnées d'analyse longitudinale élaborées par les biostatisticiens et d'autres spécialistes permettent de mener des analyses beaucoup plus nuancées que par le passé. Il faut de nombreux analystes chevronnés pour assurer l'exploitation intégrale des données recueillies dans une enquête par panel. La croissance de l'analyse secondaire (voir ci-dessous) est prometteuse comme moyen de mieux utiliser les données d'enquête par panel à l'avenir.

Enquêtes internationales. Diverses enquêtes internationales ont vu le jour depuis 25 ans, dont des enquêtes appuyées par des organismes internationaux et des enquêtes nationales indépendantes coordonnées de façon à permettre des comparaisons entre pays. Une importante percée à cet égard a été l'Enquête mondiale sur la fécondité (EMF), comportant des enquêtes menées dans 42 pays en développement et 20 pays développés au cours de la période 1974-1982. On a pu non seulement recueillir des données précieuses sur la fécondité, mais aussi, dans de nombreux pays, offrir une assistance technique pour la recherche sur les enquêtes favorisant la mise sur pied d'une infrastructure pour les sondages. Le programme d'enquêtes

démographiques et sanitaires a été lancé peu de temps après la fin de l'EMF, et des enquêtes ont été menées dans une cinquantaine de pays.

L'éducation a fait l'objet de nombreuses enquêtes internationales, y compris la Troisième étude internationale de mathématiques et des sciences (41 pays en 1995) et sa répétition (40 pays en 1999); le Programme international pour le suivi des acquis des élèves (quelque 30 pays en 2000); la deuxième *Civics in Education Study* (quelque 20 pays en 1999); la *Reading Literacy Study* de l'IEA (quelque 30 pays en 1991). L'Enquête internationale sur l'alphabétisation des adultes, qui se poursuit, permet de recueillir des données comparables sur l'alphabétisation des adultes dans plusieurs pays. Deux autres exemples sont l'enquête en grappe à indicateurs multiples (UNICEF) et la *Social Dimensions of Adjustment Integrated Survey* (Banque mondiale). Une activité connexe est la coordination, par Eurostat, des enquêtes de l'Union européenne. Un exemple de collaboration de plusieurs pays à des enquêtes est le programme international d'enquêtes sociales, programme annuel en sciences sociales qui regroupe actuellement 33 pays membres.

Les programmes d'enquêtes internationales ont pris de l'ampleur pour deux raisons distinctes. La première est l'intérêt accru manifesté pour la comparaison des résultats d'enquêtes de plusieurs pays. La deuxième est de venir en aide à des pays en développement et en transition, surtout, leur expérience des enquêtes étant limitée, afin qu'ils puissent mener des enquêtes produisant d'importantes données de planification. On peut s'attendre, pour ces mêmes raisons, à un élargissement appréciable des activités d'enquêtes internationales à l'avenir.

Liens avec les données administratives. La puissance accrue des ordinateurs et la capacité qui en résulte de mener des analyses plus perfectionnées ont suscité une demande de données plus nombreuses sur les unités échantillonnées. Les analystes veulent trouver une réponse à des questions plus complexes que ce n'était le cas par le passé, et certaines données dont ils ont besoin ne se laissent pas facilement recueillir dans le cadre d'une enquête, du moins pour le niveau de qualité exigé. Même si l'on pouvait recueillir les données, la collecte entraînerait un fardeau de réponse excessif. On a donc cherché d'autres sources de données, quitte à lier celles-ci aux réponses d'enquête. Ainsi, les dossiers d'impôt peuvent fournir de précieux profils des gains d'individus échantillonnés au cours d'une période de temps pour laquelle les répondants ne sauraient fournir les données, et les dossiers médicaux peuvent indiquer l'ampleur des dépenses médicales assumées directement par les assureurs alors qu'elle est inconnue des répondants. Ce genre de lien a été largement facilité par le nombre appréciablement accru de documents administratifs sur support électronique.

La possibilité de lier des données de documents administratifs à des données d'enquête sociale suscite un intérêt considérable depuis quelques années, et certaines enquêtes

ont établi ce genre de lien. Toutefois, il faut généralement surmonter des difficultés appréciables pour avoir accès à des données administratives, et la protection de la vie privée des répondants est une question sérieuse. Jusqu'à présent, ces considérations ont beaucoup limité le recours à des liens avec les documents administratifs dans les enquêtes-ménages. Malgré les avantages éventuels appréciables de ce genre de lien, il n'est pas clair dans quelle mesure ces obstacles pourront être contournés.

Par contre, les données administratives sont devenues un élément clé des enquêtes économiques et des recensements; dans un certain nombre de cas, elles ont remplacé les données recueillies antérieurement auprès des répondants. Il en est résulté une baisse appréciable du fardeau de réponse, une amélioration de la qualité des données, une diffusion plus rapide des résultats et une réduction des coûts.

Analyse secondaire. La puissance accrue des ordinateurs, la multiplication des enquêtes et l'accès à des données d'enquête plus détaillées ont favorisé une croissance marquée de l'analyse secondaire. Les fichiers à grande diffusion sont plus facilement accessibles, parfois dans le cadre d'archives de données, de sorte que les analystes secondaires peuvent mener leurs propres études, les données d'enquête étant ainsi analysées de façon plus approfondie. Une attention accrue doit dès lors être accordée à la protection de la vie privée des répondants; il faut s'assurer que les fichiers de données accessibles aux analystes secondaires ne portent pas atteinte à la confidentialité. Puisque l'analyse secondaire va sans doute continuer de prendre de l'ampleur, il va falloir continuer de trouver des façons de diffuser les données d'enquête tout en protégeant les répondants, sans pour autant limiter sérieusement le genre d'analyse qu'il est possible de mener.

Qualité des enquêtes. Une attention accrue est accordée aux différents aspects de la qualité des enquêtes. Depuis quelques années, divers organismes d'enquête s'intéressent à la qualité du processus d'enquête, appliquant à celui-ci les concepts de la gestion de la qualité totale. On accorde une attention plus grande que par le passé à la qualité prise dans son sens large, qui englobe l'exactitude des estimations, la pertinence, la rapidité de diffusion, l'accessibilité et la rentabilité, de même que dans son sens étroit d'exactitude. Les utilisateurs des estimations d'enquête et les responsables de l'analyse secondaire ont besoin de connaître la qualité globale des données d'enquête, y compris les erreurs d'échantillonnage, la non-réponse et la non-couverture, les erreurs de réponse et les erreurs de traitement. Même si ce besoin est reconnu depuis longtemps, les descriptions courantes de la qualité des enquêtes comportent souvent de sérieuses lacunes. Il semble qu'une attention accrue soit maintenant accordée à cette question. L'introduction de profils de qualité comprenant des rapports pleinement intégrés sur la qualité des données des enquêtes en cours est un phénomène important.

5. Conclusion

Cette dernière section porte sur quelques grandes perspectives de la recherche sur les enquêtes au cours des 10 à 20 prochaines années. La révolution informatique qui a transformé la recherche sur les enquêtes au cours des 25 dernières années se poursuit, et on peut s'attendre à d'autres modifications de plusieurs aspects de la collecte, du traitement et de l'analyse des données d'enquête. De plus, les télécommunications évoluent rapidement, et il est probable que les changements auront un effet sur les modes de collecte des données d'enquête. Il semble probable qu'à l'avenir on aura davantage recours à des plans de sondage en mode mixte, pour bénéficier de nouveaux modes lorsque les répondants y ont accès (par exemple, Internet), les modes traditionnels étant utilisés pour d'autres répondants. L'effet du mode de collecte sur les réponses d'enquête restera donc une préoccupation importante.

En général, il semble probable que la demande de données d'enquête continuera de prendre rapidement de l'ampleur à mesure que les analystes de la politique apprennent à bénéficier des données d'enquête. De plus en plus, on aura besoin d'estimations d'enquête pour de petits domaines, surtout des régions géographiques, les décideurs concevant leurs programmes en fonction de sous-groupes particuliers de la population. À l'heure actuelle, la demande de données d'enquête relève principalement des administrations centrales; à l'avenir, la demande de la part des administrations provinciales et locales pourrait s'accroître. Or le coût des enquêtes est presque le même pour les faibles populations comme pour les grandes. Les administrations locales ne pourront peut-être pas toujours se permettre le coût d'une enquête à moins que l'on ne trouve des méthodes peu coûteuses.

La principale inquiétude liée à l'avenir de la recherche sur les enquêtes est que la volonté des répondants de participer aux enquêtes continue de baisser et que le perfectionnement de la collecte des données ne réussisse pas à neutraliser cet effet. Il en résultera une baisse des taux de

réponse. Cette observation est particulièrement importante pour les enquêtes téléphoniques, dont le taux de non-réponse est déjà élevé. Une hausse appréciable du taux de non-réponse des enquêtes téléphoniques pourrait même signaler la fin de la collecte des données par téléphone pour les enquêtes-ménages.

Enfin, la prochaine décennie sera peut-être marquée par l'établissement d'une nouvelle société distincte pour les professionnels de la recherche sur les enquêtes, assurant une meilleure représentation des intérêts de tous les membres de la profession. Puisque l'échantillonnage a dominé l'évolution de la recherche sur les enquêtes au cours des premières années, celle-ci comporte des liens étroits avec les sociétés statistiques. Toutefois, ces liens se fondent surtout sur la statistique des enquêtes. Il existe aussi des liens avec les sociétés qui s'occupent de sondages, d'études de marché et de diverses spécialisations comme la sociologie et la psychologie, surtout pour ce qui est des aspects de la recherche sur les enquêtes autres que l'échantillonnage. On constate également des liens avec les sociétés d'informatique pour les personnes qui s'occupent de calculs d'enquête. Toutefois, il n'existe toujours pas de société qui cherche à réunir toutes les disciplines de la recherche sur les enquêtes. D'ici quelques années, on verra peut-être la création d'une telle société visant à favoriser les échanges entre diverses disciplines en vue de l'avancement du secteur. Un tel tournant n'éliminerait pas le besoin de maintenir les liens actuels entre les spécialistes de la recherche sur les enquêtes et les sociétés statistiques et autres. Ces spécialistes doivent se tenir au courant des principales activités de la recherche sur les enquêtes tout en suivant l'évolution de leur propres disciplines.

Remerciements

Je tiens à remercier Joe Waksberg et Dan Levine des précieuses remarques qu'ils ont formulées en vue de la préparation du présent exposé.