

Article

La qualité des enquêtes

par Lars Lyberg

Décembre 2012



Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Comment accéder à ce produit

Le produit n° 12-001-X au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de parcourir par « Ressource clé » > « Publications ».

Ce produit est aussi disponible en version imprimée standard au prix de 30 \$CAN l'exemplaire et de 58 \$CAN pour un abonnement annuel.

Les frais de livraison supplémentaires suivants s'appliquent aux envois à l'extérieur du Canada :

	Exemplaire	Abonnement annuel
États-Unis	6 \$CAN	12 \$CAN
Autres pays	10 \$CAN	20 \$CAN

Les prix ne comprennent pas les taxes sur les ventes.

La version imprimée peut être commandée par les moyens suivants :

- Téléphone (Canada et États-Unis) 1-800-267-6677
- Télécopieur (Canada et États-Unis) 1-877-287-4369
- Courriel infostats@statcan.gc.ca
- Poste
Statistique Canada
Finances
Immeuble R.-H.-Coats, 6^e étage
150, promenade Tunney's Pasture
Ottawa (Ontario) K1A 0T6
- En personne auprès des agents et librairies autorisés.

Lorsque vous signalez un changement d'adresse, veuillez nous fournir l'ancienne et la nouvelle adresse.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Notre organisme » > « Offrir des services aux Canadiens ».

Publication autorisée par le ministre responsable de
Statistique Canada

© Ministre de l'Industrie, 2012

Tous droits réservés. L'utilisation de la présente
publication est assujettie aux modalités de l'entente de
licence ouverte de Statistique Canada (<http://www.statcan.gc.ca/reference/licence-fra.html>).

This publication is also available in English.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, ses entreprises, ses administrations et les autres établissements. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0^s valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- ^p provisoire
- ^r révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- E à utiliser avec prudence
- F trop peu fiable pour être publié
- * valeur significativement différente de l'estimation pour la catégorie de référence (p<0,05)

La qualité des enquêtes

Lars Lyberg ¹

Résumé

La qualité des enquêtes est un concept multidimensionnel issu de deux démarches de développement distinctes. La première démarche suit le paradigme de l'erreur d'enquête totale, qui repose sur quatre piliers dont émanent les principes qui guident la conception de l'enquête, sa mise en œuvre, son évaluation et l'analyse des données. Nous devons concevoir les enquêtes de façon que l'erreur quadratique moyenne d'une estimation soit minimisée compte tenu du budget et d'autres contraintes. Il est important de tenir compte de toutes les sources connues d'erreur, de surveiller les principales d'entre elles durant la mise en œuvre, d'évaluer périodiquement les principales sources d'erreur et les combinaisons de ces sources après l'achèvement de l'enquête, et d'étudier les effets des erreurs sur l'analyse des données. Dans ce contexte, on peut mesurer la qualité d'une enquête par l'erreur quadratique moyenne, la contrôler par des observations faites durant la mise en œuvre et l'améliorer par des études d'évaluation. Le paradigme possède des points forts et des points faibles. L'un des points forts tient au fait que la recherche peut être définie en fonction des sources d'erreur et l'un des points faibles, au fait que la plupart des évaluations de l'erreur d'enquête totale sont incomplètes, en ce sens qu'il est impossible d'inclure les effets de toutes les sources. La deuxième démarche est influencée par des idées empruntées aux sciences de la gestion de la qualité. Ces sciences ont pour objet de permettre aux entreprises d'exceller dans la fourniture de produits et de services en se concentrant sur leurs clients et sur la concurrence. Ces idées ont eu une très grande influence sur de nombreux organismes statistiques. Elles ont notamment amené les fournisseurs de données à reconnaître qu'un produit de qualité ne peut pas être obtenu si la qualité des processus sous-jacents n'est pas suffisante et que des processus de qualité suffisante ne peuvent pas être obtenus sans une bonne qualité organisationnelle. Ces divers niveaux peuvent être contrôlés et évalués au moyen d'ententes sur le niveau de service, de sondages auprès des clients, d'analyses des paradonnées en recourant au contrôle statistique des processus et d'évaluations organisationnelles en se servant de modèles d'excellence opérationnelle ou d'autres ensembles de critères. À tous les niveaux, on peut rehausser la qualité en lançant des projets d'amélioration choisis selon des fonctions de priorité. L'objectif ultime de ces projets d'amélioration est que les processus concernés s'approchent progressivement d'un état où ils sont exempts d'erreur. Naturellement, il pourrait s'agir d'un objectif impossible à atteindre, mais auquel il faut tenter de parvenir. Il n'est pas raisonnable d'espérer obtenir des mesures continues de l'erreur d'enquête totale en se servant de l'erreur quadratique moyenne. Au lieu de cela, on peut espérer qu'une amélioration continue de la qualité par l'application des idées des sciences de la gestion ainsi que des méthodes statistiques permettra de minimiser les biais et d'autres problèmes que posent les processus d'enquête, afin que la variance devienne une approximation de l'erreur quadratique moyenne. Si nous y arrivons, nous aurons fait coïncider approximativement les deux démarches de développement.

Mots clés : Gestion de la qualité ; erreur d'enquête totale ; cadre de la qualité ; erreur quadratique moyenne ; variabilité des processus ; contrôle statistique des processus ; utilisateurs des données d'enquête.

1. Introduction

Le présent article a été rédigé en reconnaissance des apports uniques et du leadership de Joe Waksberg dans le domaine des techniques d'enquête. J'ai pris connaissance des travaux de Joe pour la première fois en lisant son article sur les erreurs de réponse dans les enquêtes sur les dépenses, rédigé en collaboration avec John Neter (Neter et Waksberg 1964). Entre autres, cet article m'a fait découvrir le phénomène cognitif appelé télescopage. Plus tard, j'ai eu l'occasion de travailler avec Joe à la préparation de la première conférence et monographie sur les méthodes d'enquête téléphonique en tant que membre du comité de rédaction (Groves, Biemer, Lyberg, Massey, Nicholls et Waksberg 1988). Nous avons également collaboré à la préparation de nombreuses conférences Morris Hansen dont les exposés ont été publiés dans le *Journal of Official Statistics* (JOS) durant mon mandat de rédacteur en chef. Joe lui-même a donné la sixième conférence, qui a été publiée dans le JOS

(Waksberg 1998). Joe était un fantastique chef de file et c'est pour moi un grand honneur d'avoir été invité à rédiger cet article sur la qualité des enquêtes, sujet qui le préoccupait beaucoup.

Bon nombre de mes amis m'ont fait part de leurs opinions ou m'ont envoyé de la documentation en prévision du présent article. Je remercie tout spécialement Paul Biemer, Dan Kasprzyk, Fritz Scheuren, Dennis Trewin et Maria Bohata de leur aide.

La qualité des enquêtes est un concept vague, quoiqu'intuitif, ayant de nombreuses significations. Dans le présent article, je discute de certaines observations qui ont trait à l'élaboration et au traitement du concept au cours des soixante-dix dernières années et, dans le cas de certains développements, il m'est même possible de remonter à des origines encore plus lointaines. Toutefois, ma discussion porte en majeure partie sur les questions qui se posent aujourd'hui dans les organismes statistiques gouvernementaux. C'est au domaine de la statistique officielle

1. Lars Lyberg, Département de statistique, Université de Stockholm, 10691 Stockholm, Suède. Courriel : Lars.Lyberg@stat.su.se.

qu'appartiennent la plupart des exemples de qualité des enquêtes que j'expose.

La présentation de l'article est la suivante. La section 2 traite du paradigme de l'erreur d'enquête totale, y compris les typologies de l'erreur, le traitement des erreurs et la conception des enquêtes en tenant compte de toutes les sources d'erreur. La section 3 porte sur les approches de gestion de la qualité qui ont eu un effet important sur les organismes d'enquête depuis le début des années 1990. Cet effet s'est manifesté par des méthodes et des approches telles que la prise en compte de l'utilisateur ou du client, la discussion des coûts et des risques dans le cadre de la recherche sur les enquêtes, et la nécessité pour les organismes de continuer à s'améliorer. La section 4 fournit des exemples de projets d'amélioration de la qualité entrepris par les organismes d'enquête. La section 5 traite des difficultés que pose la mesure, directe ou indirecte, de la qualité au moyen d'indicateurs. Est également abordée la façon dont ces mesures doivent être communiquées aux utilisateurs ou aux clients. Enfin, la section 6 offre certaines réflexions quant à la manière dont les pratiques d'enquête *doivent* évoluer afin de mieux répondre aux besoins des utilisateurs. La dernière section est réservée à la bibliographie.

2. Le paradigme de l'erreur d'enquête totale

2.1 Bref historique de l'échantillonnage

Un certain nombre d'articles décrivent l'élaboration des premières méthodes d'échantillonnage. On constate dans ces premiers travaux une reconnaissance implicite ou explicite des problèmes de qualité, même s'ils sont masqués sous des termes tels que « erreurs » et « utilité de l'enquête » (Deming 1944). Les aperçus historiques que l'on trouve, par exemple, dans Kish (1995), Fienberg et Tanur (1996), et O'Muircheartaigh (1997) insistent tous sur le fait que, jusqu'à 1950, on a assisté au plein essor de la théorie de l'échantillonnage. Dans les années 1920, l'Institut international de statistique a accepté de promouvoir les idées sur l'échantillonnage représentatif proposées par Kiear (1897) et Bowley (1913). En 1934, Neyman a publié son article historique sur la méthode représentative. Plus tard, le principe de randomisation de Fisher (1935) a été appliqué à l'échantillonnage en agriculture et Neyman (1938) a élaboré l'échantillonnage par grappes, l'estimation par le ratio et l'échantillonnage à deux phases, et introduit le concept d'intervalle de confiance. Neyman a montré que l'erreur d'échantillonnage pouvait effectivement être mesurée en calculant la variance de l'estimateur. Bill Cochran, Frank Yates, Ed Deming, Morris Hansen et bien d'autres ont perfectionné les concepts de la théorie de l'échantillonnage. Hansen a dirigé un groupe de recherche au U.S. Census

Bureau, où avait lieu à l'époque une grande partie des activités de recherche appliquée et d'élaboration de nouvelles théories. L'un des résultats remarquables des travaux du Census Bureau a été la production d'un manuel en deux volumes sur la théorie et les méthodes d'échantillonnage (Hansen, Hurwitz et Madow 1953). En fait, les progrès en théorie de l'échantillonnage étaient si importants à ce moment-là que Stephan (1948) a jugé bon de rédiger un article sur l'histoire des méthodes modernes d'échantillonnage.

Très tôt, on a reconnu qu'il pouvait exister d'autres erreurs d'enquête que celles attribuées à l'échantillonnage. Il existe des écrits sur les effets du libellé des questions, dont celui de Muscio (1917). La recherche sur la conception des questionnaires était assez intensive durant les années 1940. Mahalanobis (1946) s'est attaqué aux problèmes résultants des erreurs introduites par les enquêteurs sur le terrain chargés de recueillir les données agricoles en Inde, ce qui a donné une méthode d'estimation de ces erreurs. Cette méthode, appelée « interpénétration », peut être utilisée pour estimer ce que l'on appelle les variances corrélées introduites par les intervieweurs, les vérificateurs, les codeurs et les personnes qui supervisent ces groupes. Les sources d'erreur les plus importantes étaient certainement déjà connues autour de 1950. Deming a dressé une liste des sources d'erreur (1944) qui constitue la première typologie publiée des erreurs d'enquête, et Hansen et Hurwitz (1946) ont discuté du sous-échantillonnage des non-répondants pour essayer de fournir des estimations sans biais dans une situation présentant une non-réponse initiale. Cependant, sur le plan de la méthodologie, l'accent avait été mis jusque-là sur l'élaboration de la théorie de l'échantillonnage, ce qui est assez compréhensible. Il était en effet très important de pouvoir montrer qu'il était possible de réaliser des enquêtes en s'appuyant sur l'échantillonnage, et ce, dans diverses conditions. En 1950, il avait été démontré de manière assez satisfaisante que cela était effectivement faisable. Donc, il était temps de passer à d'autres questions et aux peaufinements.

Au début, l'emploi du terme qualité était limité avant tout au contrôle de la qualité, parfois au contrôle de la qualité des opérations d'enquête. Souvent, le contrôle de la qualité se résumait à la vérification et (ou) à l'estimation de la grandeur de l'erreur pour diverses opérations. On savait que les statistiques étaient affectées par d'autres erreurs que celles émanant de l'échantillonnage, mais la façon, liée à la qualité des processus, de réduire systématiquement ces erreurs et biais restait encore à établir (Deming 1944 ; Hansen et Steinberg 1956).

Il y a 60 ans d'ici, l'utilisateur était un joueur plutôt obscur, même si les éminents concepteurs des techniques d'enquête ne l'ignoraient pas du tout. Ainsi, Deming (1950)

soutenait que, jusqu'à ce que le but soit énoncé, il n'existait aucune bonne ou mauvaise façon d'entreprendre une enquête. Certains autres statisticiens ont fait des déclarations comparables. En fait, c'est vraiment l'utilisateur qui était caché derrière des termes tels que « problème lié au domaine spécialisé », « but de l'étude » ou « fonctions clés d'un système statistique ».

Même aujourd'hui, les concepts d'enquête et de qualité sont vagues. Comme l'on souligné Morganstein et Marker (1997), les définitions variées de la qualité nuisent aux travaux d'amélioration, de sorte que nous devons au moins essayer de faire la distinction entre les différentes définitions afin de déterminer à quoi elles servent. L'une des définitions citées le plus souvent est attribuée à Joseph Juran, à savoir que la qualité est une fonction directe de l'« adaptation à l'usage prévu ». En fait, déjà en 1944, Deming avait utilisé la phrase « fitness for purpose » (adaptation au but poursuivi), non pas pour définir la qualité, mais plutôt pour expliquer ce qui faisait la réussite d'un produit d'enquête.

Longtemps, la notion de « bonne » qualité était implicitement équivalente à une faible erreur quadratique moyenne (EQM), ce qui signifie que les données doivent être exactes et que l'exactitude d'une estimation peut être mesurée par l'EQM, qui est la somme de la variance et du carré du biais. Nous avons constaté que les statistiques fondées sur des sondages doivent aussi être utiles, ce qui a été désigné plus tard par le terme « pertinentes ». Nombre des dimensions actuelles de la qualité ne représentaient pas vraiment un sujet de préoccupation à l'époque. En outre, les utilisateurs étaient habitués à ce que la réalisation des enquêtes prenne du temps ; l'actualité des données était certes à l'ordre du jour, mais pas aussi explicitement qu'aujourd'hui. Le traitement des données d'un recensement prenait des années. Les utilisateurs étaient accoutumés à une technologie qui ne permettait d'offrir que des formes assez simples d'accessibilité. Donc, il était naturel pour les utilisateurs et les producteurs de faire en sorte avant tout que le problème statistique concorde raisonnablement avec le problème du secteur spécialisé et que l'EQM soit maintenue à un niveau acceptable. Cette EQM était, et est encore, équivalente dans de nombreux cas à la variance seulement, sans ajout d'un terme de carré du biais.

Avant de poursuivre, définissons ce qu'est une « enquête ». Une *enquête* est une étude statistique conçue pour mesurer les caractéristiques de la population afin de pouvoir estimer les paramètres de cette dernière. La proportion de chômeurs à un moment donné dans une population de personnes, ou le revenu total d'une entreprise ou d'un secteur d'activité durant une période donnée sont deux exemples de paramètres. Une enquête peut être définie comme une liste de conditions préalables (Dalenius 1985a). Selon Dalenius, une étude peut être catégorisée comme une

enquête si les conditions préalables qui suivent sont satisfaites :

1. l'étude concerne un ensemble d'objets constituant une population ;
2. la population étudiée possède une ou plusieurs propriétés mesurables ;
3. le but de l'étude est de décrire la population au moyen d'un ou de plusieurs paramètres définis en fonction des propriétés mesurables, ce qui nécessite l'observation (d'un échantillon) de la population ;
4. pour arriver à observer la population, une base de sondage est nécessaire ;
5. un échantillon d'objets est sélectionné à partir de la base de sondage conformément à un plan d'échantillonnage qui spécifie un mécanisme probabiliste et une taille d'échantillon n (où n pourrait être égal à N , la taille de la population) ;
6. des observations sont faites sur l'échantillon conformément à un procédé de mesure (c'est-à-dire une méthode de mesure et une prescription concernant son utilisation) ;
7. un processus d'estimation fondé sur les mesures est appliqué pour calculer des estimations des paramètres lorsque l'on fait une inférence au sujet de la population étudiée d'après l'échantillon.

Cette définition énumère implicitement les sources particulières d'erreur présentes dans les travaux d'enquête. Pour chaque source, il existe un certain nombre de méthodes qui en minimisent les effets, mais mesurent également leur grandeur (Biemer et Lyberg 2003 ; Groves, Fowler, Couper, Lepkowski, Singer et Tourangeau 2009).

Les écarts par rapport à la définition reflètent des défauts de qualité. En outre, les écarts de ce genre sont fréquents. Dans certains plans de sondage, les probabilités de sélection sont inconnues ou l'estimateur de la variance choisi n'est pas nécessairement celui qui convient le mieux, étant donné le plan utilisé. Le fait que ces défauts posent problème ou non dépend du but de l'enquête.

2.2 Les composantes du paradigme de l'erreur d'enquête totale

Le paradigme de l'erreur d'enquête totale est un cadre théorique utilisé pour optimiser les enquêtes en minimisant la grandeur cumulée des erreurs provenant de toutes les sources, étant donné des contraintes budgétaires. En pratique, cela signifie que nous voulons minimiser l'erreur quadratique moyenne de certaines estimations fondées sur les données d'enquête, à savoir celles que les principales parties prenantes jugent les plus importantes. L'erreur quadratique moyenne, qui est la mesure utilisée le plus fréquemment pour évaluer le travail d'enquête, est égale à la

somme des variances et des termes de biais au carré provenant de chaque source connue d'erreur. Groves et Lyberg (2010) résumant la situation du paradigme dans la pratique passée et contemporaine des enquêtes.

L'idée selon laquelle les enquêtes doivent être conçues en tenant compte de toutes les sources d'erreur émane des premiers témoins du domaine. Morris Hansen, Bill Hurwitz, Joe Waksberg, Leon Pritzker, Ed Deming et d'autres au U.S. Census Bureau, Leslie Kish à l'Université du Michigan, P.C. Mahalanobis à l'Institut statistique de l'Inde, et Tore Dalenius, à l'Université de Stockholm, étaient parmi les chefs de file de la recherche sur les enquêtes, mettant l'accent sur les erreurs et l'optimisation du plan de sondage. Ils étaient préoccupés par les limites inhérentes à la théorie de l'échantillonnage, car les erreurs non dues à l'échantillonnage risquaient de faire s'effondrer la théorie. Très pragmatiques, ils réfléchissaient beaucoup à la façon d'équilibrer les erreurs et aux coûts qu'entraîne leur traitement. Voyant des similarités entre une chaîne de montage d'usine (Deming et Geoffrey 1941) et la mise en œuvre de certains processus d'enquête, certains d'entre eux ont introduit des méthodes de contrôle tirées d'applications industrielles.

Dalenius (1967) s'est rendu compte qu'il n'existait pas encore de « formule de conception » pouvant fournir une solution optimale au problème. L'approche adoptée par Dalenius ainsi que par Hansen, Hurwitz et Pritzker (1967) consistait à minimiser tous les biais et à opter pour un scénario de variance minimale, pour que la variance devienne une approximation de l'EQM. Cela était censé se faire au moyen de schémas de vérification intense pour les productions en cours et d'études d'évaluation d'assez grande portée pour les futures productions. En 1969, inspiré par Hansen, Dalenius a présenté une communication portant sur la conception globale des enquêtes (*total survey design*), où le terme « totale » traduisait l'idée de prendre en compte toutes les sources d'erreur. Hansen, Hurwitz, Marks et Mauldin (1951), Hansen, Hurwitz et Bershad (1961), et Hansen, Hurwitz et Pritzker (1964) ont élaboré le modèle d'enquête du U.S. Census Bureau qui tenait compte de l'effet des intervieweurs, des codeurs, des vérificateurs et des chefs d'équipe, et permettait d'estimer leur contribution à l'erreur d'enquête totale. Ces schémas d'estimation, étoffés par Bailar et Dalenius (1969), consistaient en des variations de la répétition et de l'interpénétration. L'hypothèse était que l'estimation du biais était traitée par comparaison des estimations obtenues d'après les opérations ordinaires à celles obtenues au moyen des procédures privilégiées (qui ne pouvaient pas être utilisées à grande échelle pour des raisons financières, administratives ou pratiques). Aujourd'hui, ce genre d'approche est considéré comme étant la « norme de référence » (*gold standard*).

Il a été déclaré que pour bien concevoir une enquête il fallait contrôler de manière raisonnablement efficace l'erreur totale en spécifiant avec soin les procédures d'enquête, y compris des contrôles adéquats. Hansen, Deming et d'autres s'inquiétaient du coût des contrôles, mais, alors que le contrôle statistique des processus et l'échantillonnage d'acceptation avaient été mis en œuvre par un certain nombre d'organismes d'enquête, on parlait fort peu de l'amélioration continue des processus. Une part importante du travail relatif à la qualité concernait l'estimation des taux d'erreur, le contrôle des niveaux d'erreur des opérateurs individuels et la réalisation d'études d'évaluation à grande échelle qui prenaient habituellement beaucoup de temps. Les utilisateurs ne participaient pas directement au processus de conception, mais dans le système statistique fédéral des États-Unis, ils exerçaient au moins une certaine influence sur la détermination des données qui devaient être recueillies et présentées. Dalenius (1968) fournit plus de 200 références concernant les utilisateurs et les conférences à l'intention des utilisateurs associées aux produits du système statistique fédéral des États-Unis.

Bien que Hansen, Dalenius et d'autres aient été les premiers à préconiser la conception globale des enquêtes, il était rare que les utilisateurs participent directement à la détermination finale des exigences concernant l'enquête. Assez souvent, un agent, un administrateur ou un statisticien jouait le rôle de spécialiste du domaine. Il y a plusieurs décennies, c'est comme cela que nous pensions aux utilisateurs. Leurs opinions comptaient, mais ils ne participaient pas vraiment aux prises de décisions. Cependant, au fond de nous-mêmes, nous savions qu'il ne s'agissait peut-être pas d'un modèle parfait et, à la fin des années 1970, Statistics Sweden a publié une brochure interne intitulée « Que faire si un client se présente à notre porte ».

L'approche fondamentale de conception proposée par Hansen, Dalenius et d'autres comprenait un certain nombre d'étapes, dont :

- la spécification d'un objectif idéal d'enquête ;
- l'analyse de la situation de l'enquête quant aux ressources en matière de budget, de méthodologie et d'information ;
- l'élaboration d'un petit nombre d'options de plan d'enquête ;
- l'évaluation des diverses options en se basant sur les déterminations préliminaires connexes de l'EQM et des coûts ;
- le choix de l'une des options ou d'une modification de l'une d'elles, ou la décision de ne pas procéder du tout à l'enquête ;
- l'élaboration du plan d'enquête administratif, y compris l'essai de faisabilité, un système de signalisation

concernant les processus (appelé aujourd'hui parodonnées), un document de conception et un plan B.

Les vues de Kish (1965) sur la conception des enquêtes différaient légèrement. Il favorisait les applications néobayésiennes en échantillonnage et la psychométrie prônée par certains collègues à l'Université du Michigan (Ericson 1969 ; Edwards, Lindman et Savage 1963). Par exemple, Kish aimait l'idée que des estimations au jugé des biais de mesure pourraient être combinées aux variances d'échantillonnage pour construire des estimations plus réalistes de l'erreur d'enquête totale. Quant au problème d'optimisation, il pensait qu'une approche polyvalente était économiquement favorable pour les enquêtes, mais qu'il pourrait être difficile de décider sur quoi fonder le plan de sondage. Si l'on arrive à désigner une statistique principale, celle-ci peut à elle seule déterminer le plan de sondage et s'il existe un petit nombre de ces statistiques, il est possible d'opter pour un plan de sondage de compromis ; par contre, si les statistiques sont trop disparates, il pourrait n'exister aucun plan de sondage raisonnable. Kish insiste aussi sur la nécessité d'obtenir des renseignements sur le plan de sondage au moyen d'enquêtes pilotes et de prétests afin de prendre plus facilement les décisions concernant ce plan. Il a constaté que le plan de sondage et les mesures pouvaient varier considérablement selon l'environnement, tandis que l'échantillonnage changeait moins. Il pourrait s'agir de l'une des raisons pour lesquelles l'échantillonnage peut être classé facilement parmi les théories et méthodes statistiques classiques, alors qu'il est plus difficile d'insérer le processus d'enquête dans une discipline particulière (Frankel et King 1996 dans leur entrevue avec Kish).

Comme les autres ténors de la conception des enquêtes, Kish insistait sur l'importance d'un faible biais, mais appréciait le fait que la réduction du terme de biais pourrait accroître l'erreur totale. Il avait à cœur d'arriver à un équilibre raisonnable entre les diverses sources d'erreur et la façon dont les structures d'erreur variaient sous diverses options de plan de sondage. Comme Hansen et ses collègues, Kish pensait que les renseignements pertinents devaient être enregistrés simultanément durant la mise en œuvre (de nouveau nous voyons le parallèle avec les parodonnées). Hansen et ses collègues s'inquiétaient vraiment de l'application de contrôles excessifs, mais inadéquats. Ils se sont rendu compte que certains contrôles devraient peut-être être relâchés en raison des améliorations limitées qui en découlaient et que le degré d'amélioration des estimations devait être vérifié avant de procéder à tout relâchement des contrôles. Ils ont également suggéré que l'on devrait peut-être compromettre la pertinence pour obtenir des mesures contrôlables ou s'abstenir de procéder à l'enquête. Tant Hansen et ses collègues que Kish défendaient vivement l'idée de mettre fin à la pratique voulant que

l'erreur d'échantillonnage soit la seule erreur d'enquête mesurée.

Un examen de la situation actuelle porte à conclure que l'on ne dispose toujours pas d'une formule de conception des enquêtes. Il n'existe pour ainsi dire aucun manuel de planification, et la littérature sur la conception des enquêtes est par conséquent peu abondante, de même que celle sur les coûts (Groves 1989 est une exception). Aucune formule de conception n'est en vue. Depuis l'élaboration du modèle d'enquête du U.S. Census Bureau, plusieurs variantes ont fait leur apparition, certaines d'entre elles assez compliquées (Groves et Lyberg 2010). Une caractéristique commune est le fait qu'elles ont tendance à être incomplètes, c'est-à-dire qu'elles ne tiennent pas compte de toutes les sources d'erreur. Sur le plan statistique, l'attention se concentre surtout sur les composantes de la variance, en particulier la variance de l'erreur de mesure. Un certain nombre d'autres faiblesses sont associées au concept de l'erreur d'enquête totale. En premier lieu, la perspective des utilisateurs fait défaut et une vaste majorité d'entre eux ne sont pas à même de mettre en doute l'exactitude des données ni d'en discuter. Les structures et les interactions complexes des erreurs n'incitent pas les contrôles extérieurs et les contacts avec les utilisateurs ont souvent tendance à porter sur des questions moins techniques, telles que l'actualité, la comparabilité et le coût des données. Les utilisateurs ne sont pas vraiment au courant des niveaux réels d'exactitude et nous en savons fort peu quant à la façon dont ils perçoivent l'information sur les erreurs et comment y donner suite.

Comme l'a fait remarquer Biemer (2001), il existe un manque de mesures systématiques des composantes de l'EQM dans les organismes statistiques. Plusieurs bonnes raisons sont à l'origine de cette situation. À la complexité, qui a déjà été mentionnée, nous pouvons ajouter des facteurs tels que les coûts, le fait qu'il est presque impossible de publier ce genre d'information au moment où les données sont diffusées et le fait qu'il n'existe aucune mesure de l'erreur totale qui tient compte de toutes les sources d'erreur, faute d'une méthodologie appropriée ou parce qu'il est impossible d'exprimer certaines erreurs. Groves et Lyberg (2010) énumèrent certaines autres faiblesses du paradigme de l'erreur d'enquête totale. Par exemple, nous devons en savoir davantage sur l'interaction entre les variances et les biais. Il se peut qu'un accroissement de la simple variance de réponse aille de pair avec une réduction du biais de réponse, disons, quand nous comparons le mode d'interview à des options d'autoadministration du questionnaire. Récemment, West et Olson (2010) ont montré que la variance due à l'intervieweur peut résulter non seulement de l'effet individuel des intervieweurs sur les réponses recueillies dans le cadre de leurs tâches, mais aussi du fait que les intervieweurs réussissent individuellement à obtenir

la coopération de divers groupes de membres de l'échantillon.

Malgré toutes ses limites, le cadre de l'erreur d'enquête totale présente des points forts assez convaincants. Il fournit une décomposition taxonomique des erreurs, sépare la variance du biais et l'observation de la non-observation, et définit les diverses étapes du processus d'enquête. Il sert de fondement conceptuel au domaine de la méthodologie d'enquête, les sous-domaines étant définis par leur structure d'erreur connexe. Enfin, il permet de cerner les lacunes dans la littérature sur la recherche, puisque toute typologie montrera que certaines étapes et processus sont plus « populaires » que d'autres. Il suffit pour s'en convaincre de comparer les portées respectives de la littérature sur la collecte des données et de celle sur le traitement des données.

Il semble toutefois que le cadre de l'erreur d'enquête totale nécessite une extension dans des directions dont certaines avaient déjà été signalées il y a un demi-siècle. Nous avons besoin de directives afin de trouver un compromis entre la mesure de la taille des erreurs et l'obtention de processus davantage exempts d'erreur. La question que se pose Spencer (1985) est celle de savoir combien de ressources nous devrions consacrer à la mesure par opposition à l'amélioration de la qualité. Nous avons également besoin de certaines directives quant à la façon d'intégrer des notions supplémentaires dans le cadre, afin qu'il devienne un cadre de la qualité d'enquête totale plutôt qu'un cadre de l'erreur d'enquête totale (Biemer 2010). Par exemple, si l'« adaptation à l'usage prévu » est le fondement conceptuel dominant, comment pouvons-nous lancer des travaux de recherche englobant la variation de l'erreur associée à différents usages ? Cet aspect est discuté à la section suivante.

3. Principes de gestion de la qualité dans les organismes d'enquête

Au cours des années 1980 et au début des années 1990, certains organismes statistiques faisaient face à de fortes pressions financières et, dans certains cas, étaient simultanément critiqués s'ils n'accordaient pas suffisamment d'attention aux besoins des utilisateurs. Les gouvernements de la Suède, de l'Australie, de la Nouvelle-Zélande et du Canada, de même que l'administration Clinton aux États-Unis souhaitaient vivement accroître l'efficacité de leurs systèmes statistiques respectifs, ainsi que l'influence exercée par les utilisateurs sur ces systèmes. Il était naturel pour ces organismes de s'inspirer des théories et méthodes de la gestion (Drucker 1985), tout spécialement ce que l'on appelle la gestion de la qualité (Juran et Gryna 1988). Grâce à cette nouvelle littérature, il était possible d'étudier le rôle du client, les problèmes de leadership, la notion d'amélioration continue de la qualité et les divers outils susceptibles

d'aider l'organisme statistique à s'améliorer. Les travaux de Deming (1986) ont influencé particulièrement les praticiens des enquêtes, car il insistait sur le rôle des statistiques dans l'amélioration de la qualité. Il faisait valoir vigoureusement l'idée que les statisticiens doivent diriger les travaux d'amélioration, puisqu'ils ont reçu une formation leur permettant de faire la distinction entre diverses formes de variation des processus. Selon lui, trop peu de chefs de file de la statistique conseillaient la haute direction des entreprises et il voulait que des statisticiens plus proactifs deviennent ce genre de chef de file. Il avait particulièrement à cœur de développer les idées de Shewhart au sujet des cartes de contrôle comme moyen de distinguer les divers types de variation, à savoir les variations ordinaires et les variations ayant une cause spéciale. Le cycle d'amélioration de Shewhart consistant à planifier-faire-contrôler-agir (*Plan-Do-Check-Act*) faisait également partie des réflexions de Deming sur la qualité (Shewhart 1939).

Naturellement, l'existence des principes de gestion remonte à des temps anciens. Juran (1995) donne une foule d'exemples de ceux qui étaient en place, par exemple, dans l'empire romain. Le savoir-faire des artisans et un système de guildes en étaient les éléments fondamentaux. Des méthodes existaient pour choisir les matières premières et les fournisseurs. Les procédés étaient inspectés et améliorés. Les travailleurs étaient formés et motivés, et les clients obtenaient des garanties. Toutes ces caractéristiques se retrouvent encore dans les systèmes de gestion d'aujourd'hui. Les développements plus contemporains comprennent les cadres de la qualité ou les modèles d'excellence opérationnelle, tels que la gestion de la qualité totale (GTQ), les normes de l'Organisation internationale de normalisation (ISO), les critères du prix de qualité Malcolm Baldrige, le modèle d'excellence de la European Foundation for Quality Management (EFQM), les Six Sigma, les Lean Six Sigma et le tableau de bord prospectif (*Balanced Scorecard*). Ces modèles ne sont pas entièrement différents les uns des autres. Ils ont souvent en commun un ensemble de valeurs et de critères d'excellence. Ils représentent plutôt une évolution naturelle que l'on peut constater dans toutes sortes de travaux.

Donc, on a assisté à l'adoption progressive des modèles de gestion de la qualité et des stratégies de qualité dans les organismes statistiques et à une fusion avec les concepts et les idées déjà appliqués par ces organismes. Mon calendrier personnel de cette évolution est le suivant (les lecteurs sont invités à produire des ensembles différents d'événements et de dates) :

1875	Taylor introduit ce qu'il appelle la gestion scientifique.
------	--

1900 à 1930 Les idées de Taylor sont appliquées, par exemple, aux chaînes de montage chez Ford et chez Mercedes Benz.

Années 1920 Fisher commence à élaborer des théories et des méthodes concernant les plans expérimentaux.

1924 Shewhart développe la carte de contrôle.

1940 Le U.S. War Department produit un guide pour l'analyse des données sur les processus.

1944 Deming présente la première classification des erreurs d'enquête.

1944 Dodge et Romig présentent la théorie et des tableaux pour l'échantillonnage d'acceptation.

1946 Deming part au Japon.

1950 Ishikawa propose le diagramme en arêtes de poisson comme outil pour déterminer les facteurs qui ont un profond effet sur le résultat du processus.

1954 Juran part au Japon.

1960 De nombreuses entreprises lancent un programme « zéro défaut ».

1960 Le U.S. Census Bureau élabore des programmes de contrôle de la qualité.

1961 Le U.S. Census Bureau lance son modèle d'enquête.

1965-1966 Kish et Slobodan Zarkovich commencent à parler de la qualité des données plutôt que des erreurs d'enquête.

Années 1970 De nombreux organismes statistiques fournissent des lignes directrices concernant la qualité.

1975 Lancement du cadre de la gestion de la qualité totale (GQT).

1976 Adoption par un organisme statistique du premier cadre de la qualité contenant plus de dimensions que la pertinence et l'exactitude.

1987 à 1989 Lancement de la norme ISO 9000, du prix Malcolm Baldrige, de la stratégie Six Sigma et des modèles de l'EFQM.

Années 1990 De nombreux organismes statistiques commencent à travailler avec des modèles d'amélioration de la qualité et d'excellence.

1997 Publication de la monographie sur la qualité des mesures et des processus d'enquête (*Survey Measurement and Process Quality*).

1998 Mick Couper introduit le concept des « paradosés » en tant que sous-ensemble des données sur les processus.

2001 Le Leadership Group (LEG)-Qualité d'Eurostat organise la première conférence sur la gestion de la qualité en statistique officielle.

2007 Les notions d'architecture opérationnelle font leur apparition dans l'univers des enquêtes.

À partir du milieu des années 1990, les principes de gestion de la qualité ont eu un immense effet sur de nombreux organismes statistiques. Il ne s'agit pas nécessairement d'un accroissement de la qualité à tous les niveaux (personne n'a vérifié ce fait). Mais les principes se sont traduits dans la plupart des organismes par une prise de conscience de l'importance du maintien de bons contacts avec les utilisateurs et avec les clients, et une aspiration à devenir « le meilleur » ou « de niveau international ». La qualité est à l'ordre du jour.

3.1 Le concept de qualité

Au cours des dernières décennies, il est devenu évident que l'exactitude et la pertinence sont des éléments nécessaires, mais non suffisants pour évaluer la qualité des enquêtes. D'autres dimensions sont également importantes pour les utilisateurs. L'élaboration de cadres de la qualité des enquêtes s'est déroulée principalement au sein des organismes de statistique officielle et a été déclenchée par le progrès technologique rapide et d'autres développements sociétaux. Les technologies de pointe ont créé des possibilités et suscité des demandes de la part des utilisateurs au sujet de dimensions éventuelles de la qualité, telles que l'accessibilité, l'actualité et la cohérence, qui n'étaient tout simplement pas mises en relief auparavant. Les décisions prises par la société sont devenues plus complexes et de portée plus mondiale, ce qui s'est traduit par des demandes de statistiques harmonisées et comparables. Donc, des cadres de qualité permettant de faire face à toutes ces demandes étaient nécessaires. Plusieurs cadres de qualité ont été élaborés et chacun comprend un certain nombre de dimensions de la qualité. L'exactitude et la pertinence ne sont que deux de ces dimensions.

Par exemple, le cadre élaboré par l'OCDE (2011) comprend huit dimensions, à savoir la pertinence, l'exactitude, l'actualité, la crédibilité, l'accessibilité, l'intelligibilité, la cohérence et la rentabilité (tableau 1). Des cadres similaires ont été établis par Statistique Canada (Statistique Canada 2002 ; Brackstone 1999), et par Statistics Sweden (Felme, Lyberg et Olsson 1976 ; Rosén et Elvers 1999). Le système statistique fédéral des États-Unis met depuis longtemps l'accent sur l'élément d'exactitude (U.S. Federal Committee on Statistical Methodology 2001), mais il apprécie certainement d'autres dimensions. Peut-être voit-il celles-ci comme étant de nature moins statistique, mais nécessitant néanmoins une part du budget total d'enquête. Le Fonds monétaire international (FMI) a élaboré un cadre qui diffère de

ceux de l'OCDE, de l'Australian Bureau of Statistics, de Statistics Sweden et de Statistique Canada. Le cadre du FMI comprend un ensemble de conditions préalables et cinq dimensions de la qualité, à savoir l'intégrité, la rigueur méthodologique, l'exactitude et la fiabilité, l'utilité, et l'accessibilité (voir Weisman, Balyozov et Venter 2010).

Si l'exactitude est insuffisante, les autres dimensions sont sans pertinence, mais l'inverse est vrai également. Des données très exactes peuvent être inutiles si elles sont diffusées trop tard pour avoir une incidence sur les décisions importantes des utilisateurs ou si elles sont présentées de façon telle que l'utilisateur a de la difficulté à y avoir accès ou à les interpréter. En outre, les dimensions de la qualité sont souvent conflictuelles. Par conséquent, fournir un produit de qualité est un savant numéro d'équilibre dans lequel les utilisateurs informés jouent un rôle important. Des conflits existent habituellement entre l'actualité et l'exactitude, puisqu'il faut du temps pour obtenir des données exactes, grâce, par exemple, à un suivi à grande échelle des cas de non-réponse. Un autre conflit est celui qui survient entre la comparabilité et l'exactitude, puisque l'application de nouvelles méthodes plus exactes pourrait perturber les comparaisons au fil du temps (Holt et Jones 1998).

Tableau 1
Cadre de qualité de l'OCDE

Dimension	Description
Pertinence	Les statistiques sont pertinentes si les besoins de l'utilisateur sont satisfaits.
Exactitude	Degré de rapprochement entre la valeur finalement retenue et la valeur réelle, mais inconnue, dans la population.
Crédibilité	Le degré de confiance qu'ont les utilisateurs dans les produits de données en fonction de la perception qu'ils ont du fournisseur des données.
Actualité	Temps écoulé entre le moment où les données sont disponibles et le moment où a eu lieu l'événement ou le phénomène qu'elles décrivent.
Accessibilité	Facilité avec laquelle les données peuvent être localisées et consultées à l'intérieur des fonds de données.
Intelligibilité	Facilité avec laquelle l'utilisateur peut comprendre, utiliser et analyser correctement les données.
Cohérence	Reflète la mesure dans laquelle les produits de données sont reliés logiquement et mutuellement concordants.
Rentabilité	Une mesure des coûts et du fardeau imposé au fournisseur par rapport à la production.

Donc, de nombreux organismes ont adopté un concept de qualité multidimensionnel comprenant non seulement l'exactitude, mais aussi d'autres dimensions. Nous pourrions parler d'un vecteur de qualité dont les composantes varient légèrement d'un organisme à l'autre, tant en nombre qu'en contenu. Plusieurs problèmes sont associés à l'approche du vecteur de qualité.

En premier lieu, son élaboration n'a pas été précédée de communications avec les utilisateurs. Les producteurs de statistiques ont cru que les utilisateurs étaient intéressés par un ensemble particulier de dimensions, même s'il est évident que la vaste majorité d'entre eux pensent que les structures d'erreur sont trop difficiles à saisir et supposent que le producteur a la responsabilité de fournir les données les plus exactes possibles. Lorsque l'utilisateur ou le client a des exigences d'exactitude particulières, un dialogue plus approfondi peut s'établir entre eux. Selon les rares études qui ont examiné la façon dont les utilisateurs perçoivent l'information sur la qualité, les utilisateurs s'intéressent surtout aux dimensions qui sont faciles à comprendre, telles que l'actualité et les indicateurs qui paraissent simples, comme les taux de réponse. L'utilisateur veut que l'organisme statistique producteur soit crédible, ce qui se traduit par la capacité de produire des données contenant des erreurs faibles ou du moins connues, et de les livrer en temps opportun, de manière fiable et accessible. L'idée qu'il serait possible de produire une mesure de la qualité totale fondée sur des évaluations pondérées des différentes dimensions n'est pas raisonnable, même si Mirotschie (1993) soutient le contraire. Dans son article, il présente des arguments en faveur d'un ensemble normalisé d'indicateurs de la qualité et donne un exemple hypothétique d'indicateurs de la qualité de données d'indexation et calcule un indice réel (dans cet exemple, les indicateurs sont la précision, la non-réponse, la fiabilité, l'actualité et les résidus). Même s'il était possible d'élaborer un indicateur composite sous forme d'un indice, l'utilisateur voudrait savoir quels indicateurs ont contribué le plus à la valeur de l'indice. Du point de vue de l'utilisateur, la valeur de l'indice la moins favorable pourrait encore refléter une situation offrant le plus haut niveau de qualité. Il est rare qu'une faible exactitude puisse être compensée par de bonnes évaluations sur d'autres dimensions, pas même dans le cas, lors des élections, des sondages faits à la sortie de l'isoloir, pour lesquels l'actualité est indispensable. L'exactitude demeure nécessaire et il est généralement reconnu que tous les organismes dignes de confiance doivent satisfaire aux normes d'exactitude (Scheuren 2001 ; Kalton 2001 ; Brackstone 2001). Phipps et Fricker (2011) donnent un aperçu des cadres de qualité et de la littérature sur l'erreur d'enquête totale. Donc, nous pouvons convenir que la qualité des enquêtes est un concept multidimensionnel faisant intervenir plusieurs caractéristiques d'un produit ou service statistique.

3.2 Les répercussions du mouvement en faveur de la qualité sur les organismes statistiques

Simplement élargir le cadre de la qualité pour passer d'une ou deux dimensions à plusieurs d'entre elles ne suffit

pas à créer un environnement propice à la qualité. À la fin des années 1980 et au début des années 1990, de nombreux organismes statistiques se sont intéressés aux problèmes de qualité dépassant les aspects habituels de la qualité des données. Les questions concernant la satisfaction des clients, la communication avec les clients, la concurrence, la variabilité des processus, le coût de la mauvaise qualité, le gaspillage, les modèles d'excellence opérationnelle, les valeurs fondamentales, les pratiques exemplaires, l'assurance de la qualité, et l'amélioration continue de la qualité ont soudain fait partie des préoccupations quotidiennes de nombreux organismes.

Les organismes qui réussissent savent qu'il est nécessaire de s'améliorer continuellement (Kaizen) pour rester en activité et ils ont mis au point des mesures qui les aident à évoluer. Cela s'applique également aux producteurs de statistiques. Les changements qui sont censés améliorer le produit statistique sont déclenchés par les demandes des utilisateurs, par la concurrence des autres producteurs et par les valeurs des producteurs qui mettent l'accent sur l'amélioration continue en tant qu'environnement de fonctionnement général. Les mesures qui peuvent aider un organisme statistique à s'améliorer sont essentiellement les mêmes que pour les autres entreprises. Elles peuvent s'appuyer sur des modèles d'excellence opérationnelle, tels que celui de l'European Foundation for Quality Management (EFQM) (1999). Les valeurs fondamentales sur lesquelles repose le modèle de l'EFQM sont l'orientation résultats, l'orientation client, le leadership et la constance des objectifs, le management par les processus et les faits, le développement et l'implication des personnes, la formation continue, l'innovation et l'amélioration, le développement des partenariats, et la responsabilité sociale de l'organisme. Ce modèle a été adopté par le Système statistique européen

(SSE) comme outil permettant aux instituts statistiques nationaux d'atteindre le niveau voulu de qualité organisationnelle. Le concept est qu'il n'est pas possible d'arriver à un produit de bonne qualité, selon les dimensions mentionnées (ou une autre définition de la qualité du produit), si l'organisme ne met pas en place de bons processus sous-jacents. On pourrait également soutenir que le moyen le plus efficace et le plus fiable d'obtenir un produit de bonne qualité est d'utiliser des processus de bonne qualité. Si nous considérons la qualité comme un concept à trois niveaux, elle peut être visualisée comme on le présente au tableau 2.

3.2.1 Qualité du produit

Les résultats qu'il est convenu de livrer sont appelés le produit. Il peut s'agir d'estimations, de jeux de données, d'analyses, de registres, de processus normalisés ou d'autre matériel d'enquête, tel que des bases de sondage et des questionnaires. La qualité du produit correspond au concept classique de qualité utilisé pour informer les utilisateurs ou les clients de la qualité du produit ou du service. Elle peut être mesurée et contrôlée par le degré de respect des spécifications et des exigences quant aux caractéristiques du produit qui forment les dimensions de qualité d'un cadre. Les mesures de l'exactitude et les marges d'erreur entrent dans cette catégorie. Sont également pertinentes les observations en vue de déterminer si les ententes de niveau de service établies avec les clients ont été respectées. En harmonie avec les principes de gestion de la qualité, il est également assez fréquent de réaliser des sondages sur la satisfaction des utilisateurs afin de découvrir ce que ceux-ci pensent des produits et services fournis.

Tableau 2
Qualité - Concept à trois niveaux*

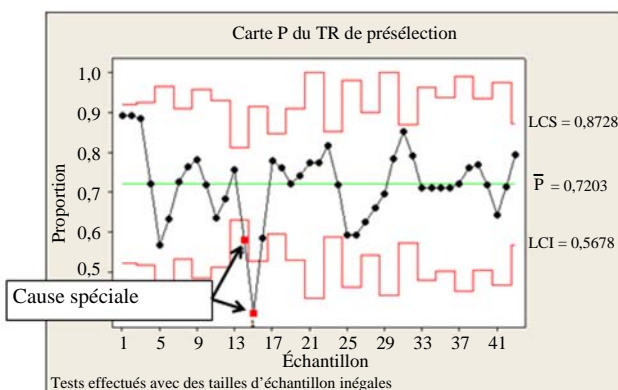
Niveau de qualité	Principaux intervenants	Instrument de contrôle	Mesures et indicateurs
Produit	Utilisateur, client	Spécifications du produit, ENS, études d'évaluation, cadres, normes	Cadres, conformité, EQM, sondage auprès des utilisateurs
Processus	Concepteur de l'enquête	CSP, cartes de contrôle, échantillonnage d'acceptation, analyse des risques, MMC, PON, paradonnées, listes de vérification, vérification	Variation au moyen de cartes de contrôle, analyse d'autres paradonnées, résultats des études d'évaluation périodique
Organisation	Organisme, propriétaire, société	Modèles d'excellence, ISO, CdP, examens, vérifications, autoévaluations	Scores, points forts et points faibles, sondages auprès des utilisateurs, sondages auprès du personnel

*ENS (Entente de niveau de service), CSP (contrôle statistique des processus), MMC (meilleures méthodes courantes), PON (procédure opérationnelle normalisée) et CdP (code de bonnes pratiques du SSE).

3.2.2 Qualité des processus

Tous les processus doivent être conçus de façon qu'ils fournissent ce qu'ils sont supposés produire. Cela signifie qu'une certaine perspective d'assurance de la qualité est nécessaire lorsque les processus sont définis. Ainsi, le processus d'interview implique qu'un certain nombre d'éléments doivent être en place pour que le processus livre ce qui est attendu. Ces éléments sont, par exemple un bon choix d'intervieweur, ainsi que la mise en place d'un programme de formation, d'un système de rémunération, ainsi que les activités de supervision et de rétroaction. Donc, nous nous employons à intégrer la qualité dans le processus par la voie de l'assurance de la qualité. Les activités de contrôle de la qualité ne sont utilisées que pour vérifier si le processus fonctionne comme il est prévu. Elles ne peuvent, à elles seules, être utilisées pour intégrer la qualité dans le processus. Cette vision des processus est discutée plus en détail à la section 4.4. La qualité des processus est mesurée et contrôlée par sélection, observation et analyses des principales variables de processus, ce que l'on appelle les données sur le processus ou par données (Morganstein et Marker 1997 ; Couper 1998 ; Lyberg et Couper 2005). La théorie et les méthodes importées du domaine du contrôle statistique des processus peuvent aider le producteur à faire la distinction entre deux types de variation, à savoir la variation ordinaire et celle ayant une cause particulière. À condition que la variation totale soit contenue entre les limites supérieure et inférieure de contrôle associées aux cartes de contrôle choisies, le processus est déclaré être sous contrôle statistique et aucune amélioration n'est vraiment possible en essayant d'ajuster les résultats individuels. Des observations qui tombent en dehors des limites de contrôle (habituellement fixées à 3 sigma), sont une indication qu'il existe une cause spéciale de variation dont il faut s'occuper afin qu'après la correction, la variation soit ramenée à celle de cause ordinaire. La carte de contrôle P qui suit illustre une situation possible :

Carte de contrôle de processus avec valeurs extrêmes



Donc, l'enchaînement des actions est le suivant. D'abord, on recherche l'origine des causes spéciales afin de pouvoir éliminer ce type de variation. Après cela, le processus ne présente plus que la variation ordinaire. Si cette variation est jugée trop grande, le processus doit être modifié. Les types de changement nécessaires sont rarement évidents au départ. En effet, plusieurs changements sont parfois nécessaires pour réduire la variation du processus. Habituellement, il faut lancer un projet d'amélioration du processus et la littérature sur la gestion de la qualité propose un certain nombre d'outils utiles pour ce genre de projet. La plupart de ces outils sont empruntés à la statistique (cartes de contrôle, expériences, analyses de régression, diagramme de Pareto, nuages de points, stratification), mais il existe aussi des outils pour déterminer les causes probables du problème (diagrammes d'enchaînement du processus, séance de remue-méninges). Une opinion répandue est que les projets d'amélioration doivent être réalisés par les personnes qui travaillent avec le processus ou par des personnes très familiarisées avec ce dernier par d'autres moyens. Parfois, nous parlons de créer une équipe d'amélioration, à laquelle participe aussi le client. Dans tout projet d'amélioration, les changements proposés doivent être testés. Quand Shewhart a développé ses premières cartes de contrôle, il a également soutenu que le travail d'amélioration doit suivre une série d'opérations qu'il a appelées *Plan-Do-Check-Act* (planifier, faire, contrôler, agir). Cette séquence nous indique que tout changement qu'il est proposé d'apporter à un processus doit être testé pour voir s'il améliore réellement le processus. Dans la négative, un autre changement est fait et les tests sont répétés. Deming a appelé ce courant de pensée le cycle de Shewhart, mais puisqu'il a passé beaucoup de temps à le promouvoir, nombreux sont ceux qui ont fini par l'appeler cycle de Deming. Les changements recherchés pourraient être une réduction de la variation du processus, une réduction des coûts, ou un accroissement de la satisfaction des clients. La méthodologie des projets d'amélioration est décrite, par exemple, dans Joiner (1994), Box et Friends (2006), Breyfogle (2003) et Deming (1986).

Un autre moyen de vérifier la qualité du processus consiste à utiliser l'échantillonnage d'acceptation (Schilling et Neubauer 2009), qui peut être appliqué à des situations où les éléments du processus peuvent être groupés par lots. Les lots sont contrôlés et, en fonction du résultat de ce contrôle, il est décidé si le lot doit être approuvé ou retravaillé. Les plans d'échantillonnage d'acceptation garantissent une qualité sortante moyenne en ce qui a trait à, disons, le taux d'erreur, mais ne comportent aucune amélioration directe de la qualité. Il s'agit d'un instrument de contrôle qui convient pour des opérations telles que le codage, la vérification et le balayage optique, et ce, uniquement quand ces processus ne sont pas vraiment en contrôle statistique. La méthode a été

vivement critiquée par Deming (1986) et d'autres, mais peut représenter le seul moyen de contrôle disponible dans des situations où le roulement du personnel est élevé et que l'on ne dispose pas de suffisamment de temps pour attendre que les processus soient stables.

Les paratonnées globales (Scheuren 2001) sont des taux d'« erreurs » de différentes sortes. Les taux de non-réponse, les taux d'erreur de codage, les taux d'erreur de balayage optique et les taux d'erreur de listage en sont des exemples. Dans le cas de certaines opérations, les taux d'erreur sont calculés en recourant à la vérification, ce qui signifie que l'opération est répétée d'une certaine façon. Il en est ainsi de l'opération de codage. Pour d'autres opérations, le calcul peut être fondé sur un schéma de classification, comme pour le calcul des taux de non-réponse. Ces paratonnées globales nous renseignent sur le processus. Il s'agit de statistiques sur les processus, c'est-à-dire des sommaires de données. Un taux de non-réponse élevé signale des problèmes dans le processus de collecte des données et un taux élevé d'erreur de codage signale des problèmes dans le processus de codage. Partant de ces données sommaires, il est parfois possible de faire la distinction entre la variation dont la cause est ordinaire et celle dont la cause est spéciale, et de décider de la mesure à prendre.

Certains processus normalisés peuvent être contrôlés au moyen de simples listes de vérification. Ces dernières sont très efficaces parce qu'il est crucial que chaque étape du processus soit accomplie, et ce, dans le bon ordre (Morganstein et Marker 1997). La préparation au décollage effectuée par les pilotes d'avion en est un exemple. Peu importe le nombre de fois qu'ils ont décollé, sans liste de vérification, le jour viendra où ils oublieront un élément. Dans le domaine de la production de statistiques, l'échantillonnage est un processus de ce genre, même si les conséquences de l'oubli d'un élément sont moins graves. Il se pourrait fort bien qu'un organisme statistique possède un processus normalisé de sélection des échantillons et qu'une liste de vérification puisse être utilisée comme directive de travail et instrument de contrôle.

Une sorte de liste de vérification peut être utilisée dans les processus plus créatifs, tels que le processus de conception globale d'enquête. Il est impossible de normaliser ce processus, mais il est possible de dresser la liste d'un certain nombre d'étapes critiques qui doivent toujours être accomplies. La liste ne nous dit pas comment les accomplir. Elle sert juste à rappeler qu'une étape particulière ne doit pas être omise ni oubliée. Morganstein et Marker (1997) discutent de ce genre de liste de vérification et les appellent (ainsi que les listes de vérification plus simples) meilleures méthodes courantes (MMC). Ils décrivent le processus d'élaboration des MMC et la façon dont ces dernières peuvent être utilisées pour réduire la variation des processus dans les

organismes statistiques. Ainsi, un organisme pourrait disposer de sept méthodes et systèmes d'imputation différents dans sa boîte à outils. Le maintien de ces sept systèmes coûte cher. Il est peu probable qu'ils soient tous aussi efficaces les uns que les autres. S'ils le sont, il n'est peut-être pas économiquement faisable de les retenir tous. Dans cette situation, une liste MMC qui décrit un plus petit nombre d'options pour l'organisme semble être une bonne idée. Elle pourrait être élaborée en créant une équipe d'amélioration comprenant les experts de l'imputation et certains clients. Les MMC sont censées être révisées lorsque de nouvelles connaissances sont acquises, ce qui implique qu'une date d'expiration est associée à chacune d'elle.

Dans un certain sens, les MMC sont naturellement des « pratiques exemplaires ». De nombreux organismes souhaitent mettre en œuvre et utiliser de telles pratiques. Morganstein et Marker offrent un processus pour élaborer ces pratiques exemplaires et les tenir à jour. Ce processus est utile pour un organisme s'il est possible de maintenir à un niveau minimum la variation de la conception des processus. Il devient alors facile de former le personnel et de modifier le processus quand il devient instable ou que de nouvelles méthodes sont mises au point. Par ailleurs, si les MMC et d'autres normes ne sont pas mises en application fermement au sein d'un organisme, leur usage ne sera pas répandu et l'investissement ne sera pas rentable.

3.2.3 Qualité organisationnelle

Les cadres sont responsables de la qualité au sens le plus large. C'est l'organisme qui assure le leadership, le perfectionnement du personnel, les outils permettant d'établir de bonnes relations avec la clientèle, les investissements et le financement. Le domaine de la gestion de la qualité nous a donné des modèles d'excellence opérationnelle qui peuvent nous aider à évaluer nos organismes statistiques de la même façon que d'autres entreprises le sont. Les deux principaux modèles d'excellence opérationnelle sont ceux du Baldrige National Quality Program et de l'European Foundation for Quality Management (EFQM).

Ces modèles consistent en une liste de critères à vérifier pour évaluer un organisme. Les sept principaux critères utilisés pour décerner le prix Malcolm Baldrige sont le leadership, la planification stratégique, l'orientation client et marché, l'information et l'analyse, l'orientation ressources humaines, la gestion des processus et les résultats opérationnels. Chaque critère comprend plusieurs sous-critères. Par exemple, l'orientation ressources humaines englobe les systèmes de travail, la formation et le perfectionnement des employés, et le bien-être et la satisfaction des employés. Les neuf critères sur lesquels s'appuie le modèle d'excellence de l'EFQM sont le leadership, les personnes, la stratégie, les partenariats et les ressources, les processus, produits et

services, les résultats personnes, les résultats clients, les résultats collectivité, et les résultats clés. Ces modèles peuvent être utilisés pour l'autoévaluation ou pour l'évaluation externe. L'organisme fournit une description de ce qui est en place en ce qui concerne chaque critère et il reçoit une note fondée sur cette description. Habituellement, les autoévaluations produisent des notes plus élevées que les évaluations externes. Il est très difficile d'obtenir une note élevée auprès d'évaluateurs externes, car les modèles sont très exigeants. Pour chaque critère, on demande à l'organisme s'il existe une bonne approche quelque part dans l'organisation. Cela est souvent le cas. La question suivante est celle de savoir dans quelle mesure cette bonne approche est répandue au sein de l'organisme. De nombreux organismes commencent à être en perte de vitesse à cette étape, car il y a peu de vrai dans la formule voulant que les bons exemples se répercutent dans toute l'organisation. Au contraire, les bonnes approches doivent habituellement être défendues énergiquement avant d'être acceptées au sein de l'organisme. La troisième question posée est celle de savoir si l'approche est évaluée périodiquement pour vérifier si elle produit les résultats attendus. C'est à ce stade que la plupart des organismes échouent. Leur stratégie consiste habituellement à épuiser une approche jusqu'à ce que les problèmes soient si importants qu'il faille la remplacer au lieu de l'ajuster. Évidemment, cette stratégie est perturbatrice et coûteuse, et ne reçoit pas une note élevée dans les évaluations de l'excellence. Le nombre maximal de points qui peut être obtenu avec ces modèles est égal à 1 000, et un gagnant obtient rarement plus de 450 à 600 points, ce qui est un indice qu'il y a matière à amélioration, même dans les organismes de calibre mondial.

Certains organismes statistiques ont utilisé des modèles d'excellence opérationnelle pour l'évaluation. L'office tchèque de statistique a été déclaré lauréat du Prix national de la qualité de la Tchécoslovaquie pour 2009 dans la catégorie Secteur public en se basant sur le modèle d'excellence de l'EFQM. Il a obtenu 464 points. Le Leadership Group-Qualité d'Eurostat a recommandé que les organismes statistiques nationaux européens utilisent le modèle de l'EFQM pour leurs travaux sur la qualité, et les organismes de la Finlande et de la Suède comptent parmi ceux qui l'ont fait. Depuis que le Leadership Group a publié son rapport en 2001 (voir Lyberg, Bergdahl, Blanc, Booleman, Grünwald, Haworth, Japac, Jones, Körner, Linden, Lundholm, Madaleno, Radermacher, Signore, Zilhao, Tzougas et van Brakel 2001), d'autres cadres et normes ont été élaborés. Le Système statistique européen a lancé son code de bonnes pratiques, qui compte un certain nombre de principes associés à des indicateurs. Pour certains de ces principes, cependant, les indicateurs constituent plutôt des éclaircissements.

La liste de principes ressemble à d'autres listes qui ont été établies par l'ONU et d'autres organisations.

Les évaluations externes sont probablement plus fiables que les évaluations internes, et ce, pour plusieurs raisons. L'une est qu'il vous est difficile de critiquer vos pairs puisque vous devrez interagir avec eux dans l'avenir ou que votre propre produit ou service sera évalué par eux dans l'avenir. Les expériences vécues à Statistics Sweden et à Statistique Canada montrent que la capacité des autoévaluations à dégager les faiblesses importantes est limitée (voir la section 5.3).

3.2.4 Certaines conséquences particulières pour les organismes statistiques

La plupart des organismes statistiques adoptent les principes de gestion de la qualité à des degrés divers et avec plus ou moins de succès. Comme l'ont fait remarquer Colledge et March (1993), il est possible d'énumérer plusieurs obstacles à la mise en œuvre de ces principes. Un organisme gouvernemental peut avoir de la difficulté à motiver son personnel au moyen de primes monétaires, puisque la façon dont l'argent des contribuables peut être dépensé fait l'objet de restrictions. La diversité des utilisateurs et des produits complique le dialogue entre le fournisseur de services et l'utilisateur, et comme il est mentionné plus haut, ni les utilisateurs ni d'ailleurs les fournisseurs des données ne sont entièrement familiarisés avec tous les biais et autres problèmes de qualité présents dans la production de statistiques. L'effet des erreurs sur les utilisations des données peut varier et est souvent inconnu. La situation se complique encore davantage du fait que, contrairement à ce que connaissent la plupart des autres entreprises, les fournisseurs des organismes statistiques ne sont pas très enthousiastes. Les fournisseurs des autres entreprises sont payés, tandis que ceux des organismes statistiques, c'est-à-dire les répondants, qui reçoivent rarement un incitatif monétaire, doivent être motivés.

Par ailleurs, les organismes statistiques ont un énorme avantage quand il s'agit d'appliquer les principes de gestion de la qualité. Un organisme statistique sait comment recueillir et analyser les données qui orientent les efforts d'amélioration. L'une des pierres angulaires des concepts de gestion de la qualité est que les décisions doivent être fondées sur des données, et souvent, les entreprises qui ne bénéficient pas de l'appui des statisticiens ne sont pas au courant des problèmes de qualité des données qui peuvent avoir des répercussions sur les décisions qu'elles prennent. Néanmoins, dans l'ensemble, un organisme statistique n'est pas différent de toute autre entreprise et il lui est fort possible d'appliquer les concepts de gestion de la qualité afin d'améliorer tous les aspects de son travail.

4. Exemples de projet d'amélioration de la qualité dans les organismes statistiques

La présente section donne des exemples de projets entrepris par les organismes statistiques en raison de l'intérêt général pour la qualité que manifeste la société.

4.1 L'erreur d'enquête totale

L'aspect peut être le plus important qu'il convient de souligner est que le domaine de la recherche et du développement portant sur la conception et la mise en œuvre des enquêtes, l'échantillonnage et les erreurs non dues à l'échantillonnage, ainsi que les effets des erreurs sur l'analyse des données, demeure florissant. L'obtention de données entachées de faibles erreurs est l'objectif principal des organismes de bonne réputation, comme en témoigne la publication régulière de manuels sur la collecte des données, l'échantillonnage, la non-réponse, la conception des questionnaires, les erreurs de mesure et les études comparatives. De nouveaux manuels traitant de sujets tels que les enquêtes auprès des entreprises, la traduction du matériel d'enquête et les paradonnées sont en cours de rédaction en vue de combler les lacunes dans ces domaines. Des revues, dont le *Journal of Official Statistics*, *Techniques d'enquête* et *Survey Practice*, sont entièrement consacrées à des sujets liés à la production de statistiques au sens large. De nombreuses autres revues, telles que le *Public Opinion Quarterly*, le *Journal of the American Statistical Association* et le *Journal of the Royal Statistical Society*, consacrent beaucoup d'espace aux méthodes d'enquête. La *Wiley series in Survey Methodology* et les conférences connexes (sur les enquêtes par panel, les méthodes d'enquête téléphoniques (deux), les erreurs de mesure, la qualité des processus, les enquêtes-entreprises, la mise à l'essai et l'évaluation des questionnaires, la collecte de données d'enquête assistée par ordinateur, la non-réponse et les enquêtes comparatives) a eu beaucoup de succès et il en est de même des ateliers continus sur la non-réponse et l'erreur d'enquête totale. Donc, les idées quant aux sources d'erreur particulières et à leur traitement ne font pas défaut. Certains domaines, tels que les erreurs de spécification, les erreurs de traitement des données et l'effet des erreurs sur l'analyse des données, sont, certes, sous-étudiés, mais, dans l'ensemble, l'élargissement des connaissances sur les erreurs d'enquête suscite un véritable intérêt. Le défi tient à la communication de ces connaissances aux personnes qui travaillent dans les organismes statistiques et à l'élaboration de principes de conception qui peuvent être appliqués pour améliorer la production des statistiques. Une fracture évidente existe entre ce qui est connu grâce à la recherche et ce qui est connu et appliqué dans les organismes statistiques. Donc, il semble nécessaire de renforcer continuellement les capacités

du personnel, en particulier parce que l'idée reçue voulant que les bons exemples se propagent comme des ondulations au sein des organismes statistiques et entre ceux-ci est un mythe. En effet, si cela se produisait vraiment, la qualité serait maintenant fantastique partout. Comme elle ne l'est pas, de nombreux organismes ont établi des programmes de formation de grande portée (Lyberg 2002).

4.2 Risque et gestion du risque

L'un des éléments de la gestion de la qualité qui a fait son entrée dans l'univers des enquêtes est le risque et la gestion de ce dernier. Eltinge (2011) parle même du risque d'enquête total (*Total Survey Risk*) comme alternative au paradigme de l'erreur d'enquête totale. L'identification et la gestion des risques est un volet important de la vérification interne moderne (Moeller 2005) et est peut-être le seul élément important absent des cadres de gestion de la qualité, tels que celui de l'EFQM. Une source d'erreur peut être jugée comme posant un plus grand risque qu'une autre et doit, par conséquent, être traitée avec plus de soin et de ressources. Par exemple, ne pas posséder de système efficace de contrôle de la divulgation statistique est considéré comme une situation très risquée. Historiquement, la divulgation illégale de données est très rare, mais lorsqu'elle a lieu, elle risque de saper toutes les tentatives ultérieures de collecte de données. Certaines décisions concernant la conception des enquêtes peuvent être considérées comme risquées. Par exemple, si nous choisissons une méthode de collecte des données qui n'est pas adaptée au sujet de l'enquête, nous pourrions obtenir des estimations qui s'écartent tellement de la vérité que les résultats seront inutiles. L'étude de comportements de caractère délicat par interview sur place ou par téléphone au lieu d'un questionnaire à remplir soi-même pourrait en être un exemple. Il existe également des risques techniques qui doivent être décelés et évalués. Ainsi, l'U.S. National Agricultural Statistical Service (Gleaton 2011) possède, comme de nombreux autres organismes, des plans de reprise après sinistre. Groves (2011) et Dillman (1996) discutent tous deux des visions différentes des risques qui pourraient émaner de la culture de production et de la culture de recherche au sein d'un organisme statistique. Le changement s'opère généralement lentement dans ces organismes, et ce, parfois, pour de bonnes raisons. Le changement pourrait aboutir à un échec, tel qu'une mise en œuvre infructueuse, des coûts importants et une réduction de la comparabilité des données. Donc, dans un certain sens, tant les producteurs que les utilisateurs des données ont tendance à hésiter à adopter les changements proposés par les chercheurs et par les innovateurs, ce qui pourrait être l'une des raisons de la lenteur avec laquelle les changements ont lieu. Il est courant de produire des mesures parallèles

pendant un certain temps afin de traiter les risques associés à la mise en œuvre d'une nouvelle méthode ou d'un nouveau système. Selon Groves (2011), la culture de la production et les utilisateurs ont eu le dernier mot au sujet de tout changement, du moins jusqu'à présent. Simultanément, l'innovation est désespérément nécessaire dans de nombreux systèmes de production et il existe des exemples d'organisations cloisonnées auxquelles il ne reste plus beaucoup de temps (avant de devoir changer), parce que les ressources pour maintenir leurs systèmes font tout simplement défaut. Donc, même en cas de résistance au changement, le manque de ressources et la concurrence feront en sorte que les organismes statistiques deviennent davantage axés sur les processus et plus efficaces. Réduire le nombre de systèmes et d'applications, et privilégier une plus grande normalisation semblent être l'une des voies d'avenir.

4.3 Le client/l'utilisateur

L'apparition des concepts de gestion de la qualité dans les organismes statistiques a rendu plus visibles les destinataires des produits et services statistiques. Les entreprises commerciales parlent toujours de leur client, tandis que les organismes gouvernementaux ont eu tendance à les appeler des utilisateurs. Quoi qu'il en soit, la prise en compte du fait que quelqu'un est censé utiliser les produits finaux ne semble pas avoir été évidente pour certains fournisseurs. Il faut admettre que l'utilisateur a été un interlocuteur depuis que l'industrie des enquêtes a vu le jour. Aux États-Unis, les conférences à l'intention des utilisateurs étaient déjà assez fréquentes il y a 50 ans (Dalenius 1968 ; Hansen et Voight 1967). Ainsi, durant six mois de 1965 à 1966, le U.S. Census Bureau a organisé 23 conférences à l'intention des utilisateurs à travers le pays et a aussi organisé des groupes consultatifs. De nombreux pays ont privilégié les contacts de nature consultative avec les utilisateurs. Les conférences à l'intention des utilisateurs ont encore lieu, mais l'apport des utilisateurs est maintenant complété par d'autres moyens, tels que des discussions publiques et des forums sur Internet. Il est rare que les utilisateurs aient participé directement à la planification et à la conception des enquêtes. Même lors des discussions au sujet de la qualité des données, les producteurs ont agi en tant que représentants des utilisateurs. Les cadres de la qualité en sont un bon exemple. Les dimensions de la qualité ont été définies en consultant très peu les utilisateurs. La littérature traitait de la façon dont ceux-ci perçoivent l'information sur la qualité est extrêmement limitée (Groves et Lyberg 2010). Qui plus est, nous ne savons pas si l'information sur la qualité que nous fournissons leur est utile (Dalenius 1985b). En fait, une supposition éclairée est que, souvent, elle ne l'est pas. Dans le cas de beaucoup d'enquêtes, les utilisateurs sont nombreux et parfois inconnus, et il n'est pas

possible de prévoir leurs besoins d'information et d'analyse. Souvent, on peut isoler un ou quelques utilisateurs principaux avec lesquels communiquer, mais bon nombre de problèmes ayant trait à la conception et à la qualité des enquêtes sont tellement compliqués qu'une grande majorité d'utilisateurs s'attendent à ce que le fournisseur de services leur livre un produit contenant la plus petite erreur possible. Hansen et Voight ont déclaré que l'exactitude devrait être d'un niveau suffisant pour éviter les problèmes d'interprétation. Aujourd'hui, il semble exister un consensus voulant que les utilisateurs recherchent des produits et des services dans lesquels ils peuvent avoir confiance, ce qui signifie que le fournisseur de services doit être crédible. La plupart des utilisateurs n'ont pas la possibilité de vérifier les niveaux d'exactitude. Les aspects dont un utilisateur moyen peut discuter sont les questions ayant trait, par exemple, à l'exactitude, à l'accessibilité et à la pertinence. Des discussions détaillées au sujet de questions techniques et de problèmes de compromis de conception entre l'exactitude et la comparabilité sont plus difficiles à obtenir.

Au cours des dernières décennies, l'utilisateur a effectivement pris plus d'importance. Certains organismes élaborent avec un utilisateur ou un client important des ententes de niveau de service qui énumèrent les exigences concernant le produit ou service final afin qu'une vérification puisse avoir lieu au moment de la livraison. De nombreux organismes qui réalisent des enquêtes auprès des entreprises ont créé des unités qui communiquent continuellement avec les entreprises les plus grandes, puisque leur participation et leur fourniture de données exactes sont absolument essentielles au processus d'estimation (Willimack, Nichols et Sudman 2002). Les grandes entreprises ne sont pas des utilisateurs au sens strict. Il s'agit de fournisseurs importants ayant souvent un intérêt dans les résultats de l'enquête. Un autre outil de communication est le sondage sur la satisfaction des clients. La valeur de ce genre de sondage est limitée en raison du phénomène d'acquiescement et de la difficulté à trouver un répondant bien informé qui est prêt à répondre. En outre, de nombreux sondages sur la satisfaction des clients s'appuient sur l'autosélection, de sorte qu'ils n'ont aucune valeur inférentielle. Les résultats de ces sondages peuvent être vus seulement comme des listes de problèmes et de préoccupations dont font part certains clients. Cette information peut évidemment être fort utile, mais elle ne convient pas pour l'estimation. De nombreux organismes d'enquête réalisent maintenant des sondages auprès des utilisateurs en continu (Ecochard, Hahn et Junker 2008).

4.4 La vue du processus

L'approche de la gestion de la qualité a de nouveau mis en relief qu'il importe d'avoir une vue du processus dans la

production des statistiques. Considérer le processus de production comme une série d'actions ou d'étapes en vue d'atteindre un but particulier qui satisfait l'utilisateur mène à un produit de bonne qualité. La qualité du processus est évaluée en déterminant dans quelle mesure chaque étape satisfait à des exigences ou à des spécifications définies. Un moyen de contrôler la qualité du processus consiste à recueillir des données sur le processus qui peuvent varier avec chaque répétition de celui-ci. Les variables du processus qu'il est intéressant de surveiller sont celles qui ont un effet important sur le résultat final du processus. Donc, afin de vérifier la stabilité et la variation d'un processus, nous avons besoin de mécanismes pour cerner les variables clés et pour recueillir et analyser des données sur ces variables. La science de la gestion de la qualité nous a donné des outils tels que le diagramme en arrêtes de poisson d'Ishikawa pour déterminer quelles pourraient être les variables clés du processus. La méthodologie de contrôle statistique des processus nous a donné des outils pour faire la distinction entre la variation dont la cause est spéciale et celle dont la cause est ordinaire et déterminer comment traiter ces deux types de variation. Habituellement, nous nous servons de cartes de contrôle qui ont été développées au départ par Shewhart (Deming 1986 ; Mudryk, Burgess et Xiao 1996) pour faire ces distinctions. Ensuite, nous recourons de nouveau à des méthodes issues de la science de la gestion de la qualité pour ajuster le processus, au besoin. Les organigrammes, ou diagrammes de flux, les diagrammes de Pareto et d'autres moyens simples permettant à l'équipe de production de repérer les causes fondamentales des problèmes en sont des exemples Juran 1988).

Les données sur les processus ont été utilisées pour contrôler les processus employés dans la production de statistiques depuis les années 1940, d'abord au U.S. Census Bureau, puis à Statistique Canada et, dans une certaine mesure, dans d'autres organismes également. Les processus habituellement vérifiés comprenaient le codage, la saisie et l'impression des données et les données sur les processus étaient principalement des taux d'erreur. Certains contrôles des processus utilisés par le U.S. Census Bureau étaient tellement compliqués et coûteux que leur valeur a été mise en doute (Lyberg 1981), surtout parce que les boucles de rétroaction qui y étaient associées étaient inefficaces et ne visaient pas toujours à déterminer les causes fondamentales des erreurs. Il était courant de blâmer les opérateurs pour les problèmes causés par les systèmes et aucun accent n'était mis sur l'amélioration continue de la qualité. À l'époque, la réflexion était davantage axée sur la vérification et la correction.

Morganstein et Marker (1997) ont conçu un plan générique d'amélioration continue du processus qui peut être appliqué à la production de statistiques. Depuis les années 1980, ils avaient travaillé dans de nombreux

organismes statistiques et constaté que, dans la plupart des cas, la réflexion au sujet de la qualité n'était pas très avancée. Ils ont fondé leur plan générique sur leurs expériences pratiques et sur les notions générales de gestion de la qualité, exposées entre autres par Juran (1988), Deming (1986), Box (1990), et Scholtes, Joiner et Streibel (1996). Ce plan compte essentiellement sept étapes :

- les caractéristiques critiques du produit sont précisées en collaboration avec l'utilisateur, en ce qui concerne les besoins généraux ainsi que les besoins plus uniques ;
- un schéma du déroulement du processus est élaboré par une équipe bien au courant. Le schéma doit comprendre la séquence des étapes du processus, les points de décision et les clients pour chaque étape ;
- les variables clés du processus sont identifiées parmi un ensemble plus grand de variables du processus ;
- la capacité de mesure est évaluée. Il est important que les décisions soient fondées sur de bonnes données et non pas simplement sur des données. Celles qui sont disponibles pourraient être inutiles. Il s'agit d'un domaine où les organismes statistiques sont avantagés par rapport à d'autres organismes. On ne doit pas tirer de conclusion au sujet de la stabilité du processus sans disposer d'information sur les erreurs de mesure. Avant tout, les données doivent permettre de quantifier l'amélioration ;
- la stabilité du processus est déterminée. Le schéma de variabilité des données sur le processus est analysé en utilisant des cartes de contrôle et d'autres outils statistiques ;
- la capacité du système est déterminée. Si la stabilité n'est pas atteinte après que la variation due à des causes spéciales a été éliminée, un effort d'amélioration est nécessaire. Des modifications peuvent être apportées au système lorsque la variation du processus est tellement grande que les spécifications, telles que les taux d'erreur minimaux ou les échéances de production, ne sont pas satisfaites. Des méthodes typiques en vue de réduire la variation sont l'élaboration et la mise en œuvre d'un nouveau programme de formation ou la mise en application d'une procédure opérationnelle normalisée. Cette dernière peut être une norme de processus, une norme relative aux meilleures méthodes courantes ou une simple liste de vérification ;
- la dernière étape du plan d'amélioration consiste à établir un système de surveillance permanente du processus. Nous ne pouvons pas nous attendre à ce que les processus restent stables au fil du temps. Pour de nombreuses raisons, une dérive s'amorce habituellement après un certain temps. Un système de surveillance facilite le suivi des nouvelles structures d'erreur, des nouvelles exigences des clients et des améliorations

possibles des méthodes et de la technologie, et permet de proposer des améliorations des processus.

Ce chapitre du livre de Morganstein et Marker a nettement influencé les travaux portant sur la qualité et la réflexion concernant les processus dans de nombreux organismes statistiques européens. L'intérêt pour ces questions s'est accru et certains organismes ont lancé leur propre système de gestion de la qualité dont l'amélioration des processus était un élément central.

Aux Joint Statistical Meetings de 1998, Mick Couper a donné un exposé sollicité sur la mesure de la qualité dans un environnement de collecte de données d'enquête assistée par ordinateur (CASIC). Il a mentionné que la nouvelle technologie produisait une foule de données secondaires susceptibles d'être utilisées pour améliorer le processus de collecte des données. Il a donné à ces données secondaires le nom de *paradonnées*, non pas dans son article, mais dans son exposé. Ce nom a été adopté très rapidement dans le monde des enquêtes et il était logique de définir la trilogie des données, des métadonnées et des *paradonnées*. Donc, nous disposons d'un terme pour les données au sujet des données (métadonnées) et un autre pour les données au sujet du processus (*paradonnées*). Les *paradonnées* sont manifestement des données sur le processus, mais très longtemps elles ont été limitées aux données au sujet du processus de collecte des données, alors que le terme utilisé dans de nombreux organismes statistiques européens était « données sur le processus » et tenaient compte de tous les processus d'enquête (Aitken, Hörngren, Jones, Lewis et Zilhao 2004). Récemment, on a assisté à un nouvel élargissement de la signification du concept. Kennickell, Mulrow et Scheuren (2009) nous rappellent ce qu'ils nomment *macro-paradonnées*, c'est-à-dire les données sur le processus global, tels que les taux de réponse, les taux de couverture, les taux de rejet au contrôle et les taux d'erreur de codage, qui ont toujours été des indicateurs de la qualité du processus dans les organismes statistiques. Lyberg et Couper (2005), Kreuter, Couper et Lyberg (2010), et Smith (2011) emploient aussi la signification plus inclusive des *paradonnées*, qui tient compte d'autres processus que la collecte des données. Il existe un risque que, comme celui de qualité, le concept de *paradonnées* soit utilisé exagérément. On trouve des exemples de discussions dans lesquelles toutes les données, sauf les estimations d'après les données d'enquête, sont considérées comme des *paradonnées*, ce qui, naturellement, n'a aucun sens.

Les *paradonnées* ont reçu un nom génial et elles sont nécessaires pour juger de la qualité du processus. Cependant, la prudence est de rigueur. On ne doit jamais recueillir des *paradonnées* qui ne sont pas reliées à la qualité du processus et il est important de savoir comment les analyser.

Parfois, les méthodes de contrôle statistique des processus peuvent être appliquées, mais parfois d'autres techniques analytiques sont nécessaires. Par exemple, pour pouvoir contrôler la falsification des données par les intervieweurs, il se pourrait que l'on doive examiner plusieurs processus simultanément, mais que la théorie et la méthodologie pour appuyer cette analyse ne soit pas directement disponible.

L'usage élargi de microdonnées qui ont trait à des enregistrements individuels, telles que les données sur les touches frappées et les enregistrements marqués d'un indicateur d'imputation, découle de l'utilisation des nouvelles technologies. Les procédures de collecte de données modernes produisent d'énormes quantités de ces types de *paradonnées*, tout comme le font aussi les systèmes de codage manuel assisté par ordinateur et les systèmes de codage entièrement automatisé, ainsi que les systèmes de balayage optique des données. Il n'est pas logique de limiter le concept à la collecte des données.

La science de la gestion de la qualité nous a appris à prévenir les problèmes concernant les processus au lieu de les corriger au moment où ils se manifestent, et nous a fait découvrir qu'il est important de faire la distinction entre différents types de variation du processus, puisqu'ils nécessitent des interventions différentes, que toute intervention ou amélioration touchant le processus doit être fondée sur de bonnes données et des méthodes d'analyse appropriées et que même des processus stables finissent par dériver, de sorte que la surveillance doit être permanente.

4.5 Normalisation et outils similaires

Un moyen de maintenir la qualité du processus sous contrôle consiste à réduire la variation en favorisant l'utilisation de normes et de documents similaires. Colledge et March (1997) discutent de quatre catégories de documents.

- Une norme est un document qui doit être respecté presque sans exception. Les écarts par rapport à la norme sont déconseillés et requièrent l'approbation de la haute direction. Des mesures correctives doivent être prises lorsqu'une norme n'est pas entièrement satisfaite. Un organisme peut obtenir une certification de conformité à une norme. Il en est ainsi des normes ISO dont quelques-unes sont pertinentes pour les organismes statistiques.
- Une politique doit être appliquée sans exception. Par exemple, un organisme peut avoir une politique concernant l'utilisation de mesures incitatives pour accroître les taux de réponse.
- Plusieurs organismes ont élaboré des lignes directrices pour différents aspects de la production de statistiques. Habituellement, les lignes directrices peuvent être outrepassées s'il y a de « bonnes » raisons de le faire.

- Une pratique recommandée est une pratique privilégiée, mais il n'est pas obligatoire d'y adhérer.

Les catégories de cette classification ne sont certes pas mutuellement exclusives, surtout si l'on tient compte également des aspects linguistiques et culturels. Par exemple, en suédois, les politiques et les lignes directrices sont conceptuellement très proches. Wikipédia, encyclopédie libre mais fondée sur un consensus, dit que les politiques décrivent des normes, tandis que les lignes directrices décrivent les pratiques exemplaires qui permettent de suivre ces normes. Cette phrase contient trois des catégories mentionnées par Colledge et March. Le mieux est probablement de traiter ces divers types de documents de la même façon. Ils visent tous à améliorer la qualité en réduisant divers types de variations et nous ne devons pas nous attarder trop sur la façon dont ils sont appelés.

Bien que les normes aient occupé une place importante dans la méthodologie d'enquête depuis longtemps, leur rôle s'est accru depuis que les organismes statistiques ont commencé à s'intéresser à la gestion de la qualité. Les premières normes, comme celles de Hansen et coll. (1967) et du U.S. Bureau of the Census (1974), étaient axées sur la discussion de la présentation des erreurs dans les données. Au U.S. Census Bureau, toutes les publications doivent informer les utilisateurs que les données sont sujettes à erreur, que l'analyse pourrait être affectée par ces erreurs et que les erreurs d'échantillonnage estimées sont plus faibles que les erreurs totales. Dans le cas des grandes enquêtes, les erreurs non dues à l'échantillonnage doivent être traitées de manière plus détaillée, contrairement à ce qui se faisait dans le passé. De nombreux autres organismes statistiques ont adopté cette façon de penser. Par exemple, les cadres de la qualité mentionnés précédemment sont des extensions qui englobent d'autres dimensions de la qualité que l'exactitude. Le Système statistique européen a élaboré et lancé successivement ce qu'on a d'abord appelé les *Model Quality Reports* (rapports modèles sur la qualité), qui sont devenus aujourd'hui simplement la *Standard for Quality Reports* (norme pour les rapports sur la qualité) (Eurostat 2009a). La norme formule à l'intention des instituts nationaux de statistique européens (noter la complexité conceptuelle) des recommandations pour la préparation de rapports sur la qualité pour une gamme « complète » de processus statistiques et de leurs produits. La norme traite des dimensions fondamentales de la qualité, à savoir la pertinence, l'exactitude, l'actualité, l'accessibilité, la cohérence et la comparabilité.

Examinons certains exemples. En ce qui concerne l'erreur de mesure, qui fait partie de la composante d'exactitude, la norme dit qu'un rapport sur la qualité doit contenir l'information suivante :

- la détection et l'évaluation générale des principaux risques en ce qui a trait à l'erreur de mesure ;
- si elles sont disponibles, les évaluations fondées sur des comparaisons à des données externes, la répétition des interviews ou des expériences ;
- l'information sur les taux de rejet durant la vérification des données ;
- les efforts déployés pour concevoir et mettre à l'essai les questionnaires, l'information sur la formation des intervieweurs et d'autres travaux sur la réduction des erreurs ;
- les questionnaires utilisés qui doivent être annexés au rapport sous une forme ou l'autre.

En ce qui concerne l'actualité, la norme dit que les rapports doivent contenir l'information suivante :

- pour les diffusions de données annuelles ou moins fréquentes : la durée moyenne de production pour chaque diffusion de données ;
- pour les diffusions de données annuelles ou plus fréquentes : le pourcentage de diffusions effectuées à temps, selon les dates de diffusion planifiées ;
- les raisons des diffusions tardives.

La norme comprend aussi des sections sur la façon de communiquer l'information au sujet des compromis entre les diverses dimensions de la qualité, l'évaluation des besoins et les perceptions des utilisateurs, le rendement et le coût, le fardeau de réponse, ainsi que la confidentialité, la transparence et la sécurité. Bien qu'elles comprennent une section sur les besoins et les perceptions des utilisateurs, ces derniers n'ont manifestement pas participé à la préparation de la norme proprement dite. Nous n'en savons toujours que fort peu sur la façon dont les utilisateurs perçoivent et utilisent l'information au sujet de la qualité. La norme est appuyée par un manuel beaucoup plus détaillé concernant les rapports sur la qualité (Eurostat 2009b) et les deux documents s'articulent autour des 15 principes énumérés dans le Code de bonnes pratiques de la statistique européenne, qui constitue le cadre de qualité fondamental pour le Système statistique européen. Les principes du Code de bonnes pratiques ont trait à l'indépendance professionnelle, le mandat pour la collecte des données, l'adéquation des ressources, l'engagement sur la qualité, le secret statistique, l'impartialité et l'objectivité, une méthodologie solide, des procédures statistiques adaptées, la limitation du fardeau imposé au répondant, le rapport coût-efficacité, la pertinence, l'exactitude et la fiabilité, l'actualité et la ponctualité, la cohérence et la comparabilité, et enfin, l'accessibilité et la clarté. Chaque principe est accompagné d'un ensemble d'indicateurs que les organismes individuels peuvent mesurer pour déterminer s'ils sont ou non en conformité avec le Code. Certains indicateurs sont vagues et de nature

très subjective, comme « l'étendue, la précision et le coût des statistiques européennes sont proportionnés aux besoins », tandis que d'autres sont plus spécifiques, tels que « un horaire standard de diffusion des statistiques européennes est porté à la connaissance du public ». Des examens par les pairs de la conformité à un ensemble limité de principes menés en utilisant une version antérieure du Code ont révélé, ce qui n'est pas surprenant, que de nombreux organismes statistiques nationaux en Europe ont de la difficulté à les respecter (Eurostat 2011a). Par conséquent, afin de faciliter la mise en œuvre du Code, on a élaboré un cadre de soutien, appelé cadre d'assurance de la qualité (CAQ) qui contient des directives plus spécifiques concernant les méthodes et les références (Eurostat 2011b). Ce cadre semble être un document fort utile, car ses références sont principalement des résumés de l'état des connaissances dans des domaines tels que l'échantillonnage, la conception de questionnaires, la vérification, et ainsi de suite qui encourage la conformité aux pratiques exemplaires courantes.

Le Code de bonnes pratiques présente de nombreuses similarités avec les Principes fondamentaux de la statistique officielle de l'ONU (de Vries 1999). Ces principes promeuvent aussi la coopération et la coordination internationales, qui constituent, en grande partie, un élément qui fait défaut dans le développement actuel de la production de statistiques (Kotz 2005). Même des pays voisins peuvent suivre des approches très différentes et posséder des niveaux de compétence en méthodologie très différents, et les différences sont parfois difficiles à expliquer. Nous savons par expérience que la collaboration en matière de développement est difficile à réaliser. Nous nous réunissons, nous parlons et nous ramenons des idées qui pourraient s'adapter à nos systèmes. En revanche, il est plus difficile de se mettre d'accord sur des approches communes. Une norme globale qui se rapporte à la production de statistiques est l'ISO 20252 – Études de marché, études sociales et d'opinion (Organisation internationale de normalisation 2006). Il s'agit d'une norme de procédure comprenant environ 500 exigences concernant les activités de recherche au sein d'un organisme. Il s'agit d'une norme minimale portant sur ce qu'il faut faire plutôt que sur la façon dont il faut faire les choses. Elle est appropriée pour les organismes qui réalisent des enquêtes et ils peuvent faire une demande de certification. En avril 2010, plus de 300 organismes du monde entier, dont la plupart étaient des entreprises de marketing, ont été certifiés. Un organisme statistique national (Uruguay) a été certifié en 2009 et Statistics Sweden prévoit obtenir la certification en 2013, mais ces organismes nationaux sont les seuls à s'être engagés sur cette voie. La norme porte sur le système de gestion de la qualité de l'organisme, ainsi que la gestion des éléments exécutifs de la recherche, la collecte

des données, la gestion et le traitement des données, ainsi que la production de rapports sur les projets de recherche (Blyth 2012).

Les normes du système statistique fédéral des États-Unis sont axées sur la composante d'exactitude. Bien qu'il ne s'agisse pas officiellement d'une norme, le U.S. Federal Committee on Statistical Methodology (2001) propose diverses méthodes pour mesurer et communiquer les sources d'erreur dans les enquêtes. En 2002, l'Office of Management and Budget (OMB) des États-Unis a publié des lignes directrices sur la qualité de l'information (OMB 2002) dont l'objectif était d'assurer et de maximiser la qualité, l'objectivité, l'utilité et l'intégrité de l'information diffusée par les organismes fédéraux. L'OMB (2006a) a également émis des normes et des lignes directrices pour les enquêtes. Elles sont bâties d'une manière classique. Vient d'abord une norme telle que « les taux de réponse doivent être calculés en utilisant des formules normalisées pour mesurer la proportion de l'échantillon admissible qui est représentée par les unités répondantes dans chaque étude, à titre d'indicateur du biais de non-réponse possible. » Cette norme est suivie d'un certain nombre de lignes directrices indiquant comment faire les calculs nécessaires, tandis que la dernière de ces lignes directrices précise que « si le taux de non-réponse globale dépasse 20 %, une analyse du biais de non-réponse doit être effectuée pour voir si les données manquent entièrement au hasard. » Comme dans le cas des normes du SSE, les lignes directrices de l'OMB sont complétées par un document de soutien (OMB 2006b) pour faciliter le respect des normes.

Dans le système statistique fédéral décentralisé des États-Unis, la plupart des organismes ont produit des documents dans lesquels sont adaptées les lignes directrices de l'OMB. Par exemple, l'U.S. Census Bureau possède ses propres normes de qualité statistique dont le niveau de détails techniques est plus élevé que celui des documents de l'OMB. Chaque norme est décrite au moyen d'exigences et de sous-exigences, et le document fournit souvent des exemples très spécifiques d'études qui peuvent être réalisées. Le National Center for Health Statistics, le National Center for Education Statistics, et l'Energy Information Administration sont d'autres exemples d'organismes américains dotés de normes concernant la qualité de l'information diffusée. Toutes ces normes peuvent être téléchargées à partir des sites Web de ces organismes.

Statistique Canada a émis des lignes directrices concernant la qualité depuis 1985. Elles sont similaires à celles du SSE puisqu'elles ne se limitent pas à mettre l'accent sur l'exactitude. Toutefois, elles sont nettement plus détaillées et contiennent un grand nombre de références. Une caractéristique particulière est que, pour certains processus, les lignes directrices prescrivent l'utilisation du contrôle

statistique des processus. Aucun autre organisme ne semble le faire.

L'édition la plus récente de ces lignes directrices est donnée dans Statistique Canada (2009).

De nombreux autres organismes statistiques de par le monde possèdent leurs propres normes de qualité. Elles sont parfois décrites comme des lignes directrices ou des normes et parfois, comme des systèmes de soutien opérationnel ou des cadres d'assurance de la qualité. Quoi qu'il en soit, le contenu et le style varient d'un organisme à l'autre, mais il faut que la variation soit gérable. Il devrait être possible d'arriver mondialement à un plus haut degré de normalisation, puisque cela s'est fait dans d'autres domaines, tels que les voyages aériens. Apted, Carruthers, Lee, Oehm et Yu (2011) discutent des divers moyens d'industrialiser le processus de production de statistiques à l'Australian Bureau of Statistics.

La question est de savoir si des normes internationales amélioreraient la qualité des enquêtes en général. Certains domaines dans lesquels des normes seraient avantageuses comprennent le calcul des indicateurs de qualité utilisés fréquemment, tels que les taux d'erreur et les effets de plan, ainsi que les pratiques exemplaires pour la traduction du matériel d'enquête, le traitement des enquêtés ne parlant pas la langue du pays, et la pondération pour tenir compte de la non-réponse. Il ne faut pas oublier que, quand une norme est émise, elle doit être mise à jour continuellement et qu'il est bien connu qu'elles sont parfois difficiles à appliquer. Si les normes sont exhaustives, le praticien peut se sentir écrasé et, par conséquent, les ignorer en grande partie, à moins que leur application ne soit rendue obligatoire et vérifiée.

4.6 Modèles de processus opérationnels statistiques

Au cours des dernières années, des concepts tels que les modèles de processus opérationnels et l'architecture opérationnelle ont été intégrés par certains organismes statistiques dans les travaux concernant la qualité. Afin de rendre les processus de production plus efficaces et plus souples, on peut les percevoir comme faisant partie d'un modèle d'architecture opérationnelle (Reedman et Julien 2010). Dans le domaine de la production statistique, un modèle générique du processus de production statistique est élaboré conjointement par la CEE-ONU, Eurostat et l'OCDE. Tout remaniement de système doit être dicté par les demandes des clients, les évaluations des risques et les nouveaux développements. Les principes architecturaux qui sous-tendent cette école de pensée sont résumés dans Doherty (2010), qui discute du renouvellement de l'architecture à Statistique Canada.

Voici quelques-uns des principes :

- La prise de décisions doit être optimale à l'échelle de l'organisme, ce qui implique la centralisation de l'informatique, du soutien méthodologique et du traitement des données.
- L'utilisation de services intégrés, tels que la collecte, la saisie et la diffusion des données, doit être optimisée.
- La réutilisation doit être maximisée et prévoir le plus petit nombre possible de processus opérationnels distincts et le plus petit nombre possible de systèmes informatiques.
- La boîte à outils de l'organisme doit être réduite au minimum.
- Le personnel doit bien connaître les outils et les systèmes.
- Les reprises, telles que les vérifications répétées, doivent être éliminées.
- Les activités de base doivent être le point de concentration, et le travail lié au processus de soutien doit être externalisé.
- Les activités de développement doivent être séparées des opérations continues.
- La collecte électronique des données doit être considérée comme le mode initial.
- Les obstacles structurels, tels que le chevauchement ou le manque de clarté des mandats, doivent être éliminés.

Ces principes sont semblables à ceux que nous avons dégagés lorsque nous avons appliqué les principes de gestion de la qualité tirés des divers cadres et modèles d'excellence décrits plus haut. Les principes constituent une évolution de la décentralisation vers un mode de pensée plus intégré. De nombreux organismes statistiques sont conscients que le cloisonnement est une chose du passé et que le passage à une plus grande centralisation est nécessaire.

5. Mesure de la qualité

Donc, la qualité est un concept multidimensionnel et sa mesure est une tâche compliquée. Nous avons noté que la qualité des enquêtes peut être considérée comme un concept tridimensionnel associé au produit final, aux processus sous-jacents qui mènent au produit et à l'organisation qui fournit les moyens d'exécuter les processus et de livrer le produit ou le service avec succès. Il existe essentiellement deux façons de mesurer la qualité. L'une est l'estimation directe de l'erreur d'enquête totale ou de certaines composantes de cette dernière. L'autre consiste à mesurer des indicateurs de la qualité dans l'espoir qu'ils reflètent effectivement le concept proprement dit.

5.1 Estimations directes de l'erreur d'enquête totale

Les décompositions existantes de l'erreur quadratique moyenne décrite, par exemple, dans Hansen et coll. (1964), Fellegi (1964), Anderson, Kasper et Frankel (1979), Biemer et Lyberg (2003), Weisberg (2005), et Groves et coll. (2009) sont toutes incomplètes en ce sens qu'elles ne tiennent pas compte de toutes les sources d'erreur. Il est rarement possible de calculer directement l'EQM dans les situations pratiques d'enquête, parce que ce calcul nécessite en général une estimation des paramètres qui est essentiellement exempte d'erreur. Toutefois, il est possible d'obtenir une deuxième meilleure estimation de la vraie valeur des paramètres si des ressources sont disponibles pour recueillir des données par une méthode considérée comme la « norme de référence », mais qui n'est ni de coût abordable ni pratique dans des conditions normales d'enquête. Il s'agit de la méthode classique d'évaluation lorsque la valeur vraie des paramètres peut être définie de manière unique. Les méthodes considérées comme la norme de référence sont rarement exemptes d'erreur, mais elles peuvent, à des degrés variables, fournir de meilleures estimations, et l'écart entre l'estimation ordinaire et l'estimation de référence peut servir d'estimation du biais, méthode qui est utilisée dans les enquêtes postcensitaires (Nations Unies 2010). Souvent, l'évaluation porte sur une composante particulière de l'erreur, telle que le sous-dénombrement au recensement, le biais de non-réponse, la variance due à l'intervieweur ou la simple variance de réponse, puisque nous voulons de l'information non pas sur l'erreur d'enquête totale proprement dite, mais plutôt sur la contribution relative de la composante à l'erreur totale d'enquête afin de pouvoir cerner les causes fondamentales des problèmes et d'améliorer les processus pertinents. Les grandes études d'évaluation sont très rares, parce qu'elles sont exigeantes et que leur valeur est parfois mise en doute (Nations Unies 2010). Par ailleurs, des études d'évaluation régulières à plus petite échelle sont nécessaires pour obtenir des indices quant aux problèmes opérationnels et méthodologiques.

5.2 Indicateurs de qualité

La communication continue de l'erreur d'enquête totale est une tâche gigantesque qu'aucun organisme d'enquête n'entreprend. Ils fournissent plutôt des indicateurs ou des déclarations concernant la qualité. Par exemple, selon le manuel de production des rapports sur la qualité d'Eurostat (2009a), il convient de mesurer les indicateurs suivants :

- coefficient de variation ;
- taux de surcouverture ;
- taux de rejets à la vérification ;
- taux de réponses totales ;
- taux de réponses partielles ;

- taux d'imputation ;
- nombre d'erreurs ;
- ampleur moyenne des révisions.

Le thème commun ici est que ces éléments sommaires des paratonnées sont des indicateurs qui peuvent être calculés sans effectuer d'études spéciales. Le jeu d'indicateurs qui peuvent être calculés directement d'après les données d'enquête est, par définition, assez limité et leur valeur est douteuse. Par exemple, inclure la surcouverture, mais non la sous-couverture, simplement parce que la première peut être calculée directement d'après les données disponibles, n'est pas logique. C'est la sous-couverture qui pose le problème de couverture le plus important dans les enquêtes. Le manuel prescrit certes que le producteur évalue le biais possible (tant son signe que sa grandeur), mais il ne décrit pas clairement comment cela doit se faire. Il est demandé au producteur d'inclure les résultats des évaluations et des contrôles de la qualité, si cette information existe aussi. Les mesures du niveau d'effort pour des processus tels que la conception des questionnaires et la formation des codeurs seraient les bienvenues. Il n'existe aucun format normalisé de présentation de cette information qualitative et quantitative. De toute façon, la liste d'indicateurs clés est très restreinte si on la compare à la liste complète des principales sources d'erreur, et il est difficile de voir comment ces indicateurs sont perçus par les utilisateurs et comment ils peuvent être utilisés par le producteur pour améliorer le processus.

Le producteur a besoin d'une liste plus complète d'indicateurs pour pouvoir mesurer ou évaluer divers niveaux de qualité pour s'assurer que la mise en œuvre du plan d'enquête est maîtrisée ou être capable de mettre sur pied un projet d'amélioration de la qualité. Le plan d'enquête initial doit être modifié ou adapté durant la mise en œuvre afin de contrôler les coûts et de maximiser la qualité. Biemer (2010) discute de quatre stratégies de réduction des coûts et des erreurs en temps réel, à savoir l'amélioration continue de la qualité (ACQ), la collecte adaptative (Groves et Heeringa 2006), les Six Sigma (Breyfogle 2003), ainsi que la conception et la mise en œuvre totalement adaptatives.

Pour appliquer la stratégie d'amélioration continue de la qualité, il faut déterminer quelles sont les variables clés du processus ainsi que les caractéristiques de ce dernier qui sont essentielles à la qualité. Pour chaque caractéristique essentielle à la qualité, il faut élaborer des mesures fiables, en temps réel, du coût et de la qualité. Les mesures sont surveillées en permanence durant le processus et des interventions ont lieu afin de s'assurer que les coûts et la qualité demeurent dans les limites acceptables. La stratégie de collecte adaptative a été conçue pour réduire le biais de non-réponse dans les entrevues en personne. Elle comprend

trois phases. Durant la phase expérimentale, quelques options de plan de collecte sont mises à l'essai (par exemple, en ce qui concerne le niveau des mesures incitatives). Durant la phase principale de collecte de données, l'option choisie durant la phase expérimentale est mise en œuvre et se poursuit jusqu'à ce que soit atteinte la limite de capacité. Durant la phase de suivi des cas de non-réponse, des méthodes spéciales sont mises en œuvre pour réduire le biais de non-réponse et pour contrôler les coûts de la collecte des données. Ces méthodes comprennent le scénario d'échantillonnage double de Hansen-Hurwitz, l'augmentation des mesures incitatives et l'utilisation d'intervieweurs plus expérimentés. De nouveau, les efforts se poursuivent jusqu'à ce qu'une réduction supplémentaire du biais de non-réponse ne soit plus rentable. Le modèle des Six Sigma est le modèle d'excellence opérationnelle le plus élaboré, puisqu'il s'appuie fortement sur des méthodes statistiques. Il contient un jeu important de techniques et d'outils qui peuvent être utilisés pour contrôler et améliorer les processus. La conception et la mise en œuvre entièrement adaptatives combinent les caractéristiques de contrôle de l'amélioration continue de la qualité, de la collecte adaptative et du modèle Six Sigma, afin de surveiller simultanément les multiples sources d'erreur. Biemer et Lyberg (2012) donnent plusieurs exemples de caractéristiques essentielles à la qualité et de mesures pour divers processus d'enquête. Par exemple, dans le cas du processus de mesure, les attributs qui sont des caractéristiques essentielles à la qualité pourraient inclure les aptitudes à repérer et à corriger les questions d'enquête qui posent problème, à déceler et à contrôler les erreurs de réponse, et à minimiser les biais et les variances dus aux intervieweurs. Les mesures correspondantes pourraient inclure le nombre d'éléments de données manquants par question, les taux de refus selon la taille de l'entreprise, les résultats des mesures répétées, les contrôles douteux effectivement modifiés, et les résultats du travail sur le terrain par intervieweur. Les mesures peuvent être analysées en utilisant les méthodes de contrôle statistique des processus ou d'analyse de la variance. Différentes mesures connexes peuvent être affichées simultanément sous forme d'un tableau de bord. Par exemple, si l'une des caractéristiques essentielles à la qualité est la capacité de découvrir les intervieweurs qui trichent, nous pourrions créer un tableau de bord montrant la durée moyenne d'interview par intervieweur et la distribution de certaines caractéristiques de l'échantillon de nature sensible, également par intervieweur.

5.3 Autoévaluations et vérifications

Les principes de gestion de la qualité ont mené à l'introduction des concepts d'autoévaluation et de vérification dans la production de statistiques. Nous souhaitons

vivement savoir ce que les utilisateurs, les clients, les propriétaires et d'autres parties prenantes pensent des produits et services fournis par l'organisme statistique. Un certain nombre d'outils sont disponibles pour ce genre d'évaluation. Nous avons déjà mentionné le sondage sur la satisfaction des clients. Les autres outils comprennent les sondages auprès des employés, les vérifications internes et les vérifications externes. Les sondages auprès des clients peuvent jeter de la lumière sur ce que les utilisateurs pensent des produits et services qui leur sont fournis. Ils peuvent servir à déterminer les besoins des utilisateurs et à cerner les caractéristiques du produit qui importent vraiment. Une autre série de questions pourraient avoir trait à l'image de l'organisme et à sa comparaison à celle d'autres organismes, qu'il s'agisse ou non de concurrents. Les sondages sur la satisfaction des clients sont très fréquents dans notre société. Souvent, il est impossible de les utiliser pour faire des inférences au sujet de la population cible d'utilisateurs, à cause de défauts méthodologiques et conceptuels. L'abondance de sondages sur la satisfaction, développés et mis en œuvre par des personnes ne possédant aucune formation officielle en méthodes d'enquête, contribue à l'accueil tiède qui leur est réservé dans des situations plus sérieuses donnant lieu à des erreurs de non-réponse et de mesure. Ainsi, le sondage sur la satisfaction des utilisateurs mené en 2007 par Eurostat comprenait deux sondages distincts. L'un, lancé sur la page Web d'Eurostat, avait pour population cible 3 800 utilisateurs inscrits. Seuls les utilisateurs inscrits qui sont entrés dans le site Web durant la période de collecte des données ont été exposés à la demande de participation au sondage, ce qui a donné un taux de réponse d'environ 5 %. Le second sondage, réalisé par courrier électronique, a été envoyé à plusieurs utilisateurs importants identifiés par Eurostat. Cet environnement plus contrôlé a produit un taux de réponse de 28 %. Ces sondages posent aussi des problèmes pour ce qui est d'identifier le répondant le plus approprié. Le choix du « mauvais » répondant au sein d'un organisme aboutira certainement à des résultats non éclairés et trompeurs.

Le type le plus simple d'autoévaluation est le questionnaire ou la liste de vérification remplie par le gestionnaire de l'enquête. Un exemple est offert par Statistics New Zealand. Il s'agit d'une liste de vérification comprenant un certain nombre d'indicateurs ou de déclarations, tels que « les besoins d'information sont évalués régulièrement en consultant les utilisateurs », « documentation utile et accessible », « production et surveillance régulières d'indicateurs de l'exactitude » et « respect des normes de présentation ». Le gestionnaire doit répondre par oui ou par non à chaque question et faire un commentaire s'il le juge nécessaire. Statistics Sweden utilisait un système similaire dont l'une des questions était « Comparativement à l'année dernière, la qualité globale de votre produit s'est-elle améliorée, a-t-elle






diminué ou est-elle restée la même ? ». Lorsque les résultats ont été compilés pour ces trois catégories pour l'ensemble de l'organisme, il s'est avéré qu'une très faible proportion de gestionnaires avait déclaré une baisse de qualité, une proportion un peu plus élevée, une amélioration de la qualité, tandis qu'une vaste proportion avait déclaré qu'il n'y avait pas eu de changement. Les gestionnaires n'avaient tout simplement pas les moyens appropriés d'évaluer la qualité globale. En outre, des quantificateurs vagues, tels que « régulièrement », « utile » et « respect des normes », sont une invitation à fournir des évaluations généreuses. De surcroît, la plupart des gestionnaires ne veulent pas faire mauvaise impression et le *statu quo* devient l'échappatoire parfaite. Statistics Sweden a fini par abandonner ce système d'évaluation. Il est possible d'accroître la valeur de ces évaluations en posant des questions supplémentaires pour obtenir des renseignements détaillés sur la façon dont les travaux liés à la qualité ont été menés et à quel moment. Certains organismes recourent à des équipes internes qui vérifient les produits importants. Julien et Royce (2007) décrivent une vérification de la qualité de neuf produits menée par Statistique Canada afin de repérer les points faibles et leurs causes fondamentales, ainsi que pour dégager les pratiques exemplaires. Des équipes d'examen constituées de gestionnaires adjoints ont été créées afin que chaque examinateur passe en revue trois programmes différents. Le principal point faible d'une telle approche est l'élément interne proprement dit. Chaque examinateur sait que son tour d'être soumis à un examen viendra tôt ou tard, ce qui risque de le freiner. Le problème est également interne en ce sens que les utilisateurs ne sont pas explicitement présents durant le processus d'examen. Toutefois, dans son programme général de vérification de la gestion de la qualité des données, Statistique Canada insiste beaucoup sur son système de liaison avec les utilisateurs (Julien et Born 2006), qui est l'un des cinq systèmes formant le cadre d'assurance de la qualité de l'organisme, les autres étant la planification intégrée, les méthodes et les normes, la diffusion et la production de rapports sur les programmes.

Une autre variante de l'autoévaluation est celle où elle précède une vérification externe. Statistics Netherlands (1997) décrit comment le Service des méthodes statistiques a été évalué par son personnel. L'évaluation a produit une liste de points faibles et de points forts qui a ensuite été examinée par une équipe externe. Habituellement, une vérification externe s'appuie pour l'évaluation sur certaines références, telles qu'un ensemble de règles, une norme ou un code de bonnes pratiques. La vérification aboutit alors à un certain nombre de recommandations concernant l'organisme ou le produit ou service en question.

Récemment, Statistics Sweden a élaboré un système général d'évaluation de l'erreur d'enquête totale. Le

ministère des Finances de la Suède souhaite que les résultats des évaluations de la qualité permettent de suivre les améliorations de la qualité au fil du temps. Comme il faut évaluer la qualité d'un grand nombre d'enquêtes, de registres administratifs et d'autres programmes de l'organisme, il est nécessaire de disposer de certains indicateurs qui peuvent servir de mesures indirectes remplaçant les mesures réelles de la qualité. Parallèlement, le processus d'évaluation doit être complet, la communication des résultats doit être simple et les résultats doivent être crédibles. Pour chacune des sources d'erreur – spécification, base de sondage, non-réponse, mesure, traitement des données, échantillonnage, modélisation/estimation et révision –, huit produits clés ont été notés chacun comme étant mauvais, passable, bon, très bon ou excellent en ce qui concerne cinq critères. Ceux-ci étaient la connaissance des risques, la communication avec les utilisateurs, le respect des normes et des pratiques exemplaires, l'expertise disponible, et les réalisations en regard des plan d'atténuation des risques et/ou d'amélioration. Les lignes directrices de notation variaient selon le critère. Voici celles appliquées pour la connaissance des risques :

Exemple de lignes directrices de notation – Connaissance des risques

Mauvais 	Passable 	Bon 	Très bon 	Excellent 
La documentation interne sur le programme ne mentionne pas la source d'erreur comme un facteur de risque possible pour l'exactitude du produit.	La documentation interne sur le programme mentionne la source d'erreur comme un facteur de risque possible pour la qualité des données.	Un certain effort a été fait pour évaluer l'effet possible de la source d'erreur sur la qualité des données.	Des études ont été menées pour estimer les composantes pertinentes du biais et de la variance associés à la source d'erreur et elles sont bien décrites.	Il existe un programme permanent de recherche en vue d'évaluer toutes les composantes pertinentes de l'EQM associées à la source d'erreur, et leur incidence sur l'analyse des données. Le programme est bien conçu et axé sur les éléments appropriés, et fournit l'information requise pour faire face aux risques dus à cette source d'erreur.
	Mais : Aucun effort n'a été fait ou très peu d'effort a été fait pour évaluer ces risques.	Mais : Les évaluations n'ont pris en considération que des mesures indirectes (par exemple, les taux d'erreur) de l'effet, sans aucune évaluation des composantes de l'EQM.	Mais : Les études n'ont pas exploré les conséquences des erreurs sur divers types d'analyse des données, y compris les analyses de sous-groupes et de tendances et les analyses multivariées.	

Le processus d'évaluation a débuté par une auto-évaluation de chacun des huit produits clés. Les rapports de ces autoévaluations et d'autres documents pertinents ont été étudiés par des examinateurs externes qui ont ensuite rencontré les responsables des produits et leurs employés pour discuter des processus de production. Ensuite, les examinateurs ont présenté des évaluations détaillées et ont attribué une note à chaque produit. La procédure a permis de cerner d'importants domaines à améliorer, non seulement pour chaque produit, mais aussi pour l'ensemble des produits. Ce premier cycle d'évaluation a indiqué que l'erreur de mesure posait problème pour presque tous les produits clés. Comme toute autre approche de mesure ou d'indication de l'erreur d'enquête totale, celle-ci ne reflète pas vraiment l'erreur quadratique moyenne totale. Elle nécessite une description approfondie des processus et des améliorations apportées, et elle dépend fortement des compétences et des connaissances des examinateurs externes. Cette étude est présentée dans Biemer, Trewin, Japac, Bergdahl et Pettersson (2012).

5.4 Profils de qualité

Dans le cas des enquêtes continues, il est possible d'élaborer des profils de qualité. Ce genre de document contient tout ce que l'on sait de la qualité d'une enquête continue ou d'un autre produit statistique assemblé au cours de plusieurs années. Les profils de qualité n'existent que pour quelques grandes enquêtes, qui sont toutes sauf une, réalisées aux États-Unis, à savoir la Current Population Survey (Brooks et Bailar 1978), la Survey of Income and Program Participation (Jabine, King et Petroni 1990 ; Kalton, Winglee et Jabine 1998), la Schools and Staffing Survey (Kalton, Winglee, Krawchuk et Levine 2000), et l'American Housing Survey (Chakrabarty et Torres 1996). Fait exception la British Household Panel Survey (Lynn 2003). Le principal problème que pose un profil de qualité est son manque d'actualité, puisqu'il s'agit d'une compilation des résultats d'études de la qualité qui prennent souvent beaucoup de temps. L'objectif du profil de qualité est de cerner les domaines où ils existent des lacunes dans les connaissances sur les erreurs, afin de pouvoir apporter des améliorations. Kasprzyk et Kalton (2001) ainsi que Doyle et Clark (2001) passent en revue l'utilisation des profils de qualité aux États-Unis.

6. Et maintenant, où allons-nous ?

Les notions de gestion de la qualité ont été influentes dans de nombreux organismes statistiques. Des concepts tels que le leadership, la culture de la qualité, la prévention des problèmes, la clientèle, la concurrence, l'évaluation des risques, la réflexion au sujet du processus, l'amélioration,

l'excellence opérationnelle et l'architecture opérationnelle sont des sujets abordés de plus en plus fréquemment par les dirigeants des organismes d'enquête, par exemple, Trewin (2001), Pink (2010), Fellegi (1996), Brackstone (1999), deVries (1999), Groves (2011), et Bohata (2011). Le monde des enquêtes semble s'engager dans une direction où la production de statistiques devient rationnelle et rentable, mais l'évolution est lente. Certains organismes ont commencé à se servir d'un modèle de gestion de la qualité à des fins d'autoévaluation et d'orientation. Le modèle d'excellence de l'EFQM est celui qui est recommandé aux instituts nationaux de statistique qui font partie du Système statistique européen et deux d'entre eux, ceux de la République tchèque et de la Finlande, ont même posé leur candidature pour l'obtention du Prix national de l'EFQM de leur pays. Certaines entreprises de marketing sont certifiées d'après la norme ISO 9001 de management de la qualité et d'autres sont certifiées d'après la norme ISO 20252 concernant les études de marché et les études sociales et d'opinion. Ce développement devrait aboutir à des améliorations de la qualité, mais nous ne pourrions pas vraiment en être certains tant que nous ne commencerons pas à recueillir des données pertinentes. Cependant, une chose est sûre. Certains clients préfèrent les fournisseurs de services qui sont certifiés, qui ont gagné des prix ou qui peuvent donner la preuve que leur travail est conforme à un cadre ou à un modèle de qualité. Rares sont les clients qui jugeraient qu'un tel attribut est négatif.

Les marges d'erreur que nous associons aux estimations sont habituellement trop étroites, puisqu'elles n'englobent pas toutes les sources de variation. Les estimations ponctuelles peuvent se situer hors des limites à cause des biais. Idéalement, il serait utile de pouvoir produire des estimations de l'erreur d'enquête totale plutôt que celles produites aujourd'hui. Toutefois, ce genre de progrès n'est pas réaliste. Nous ne sommes pas en mesure de produire ce genre d'estimations, même à l'occasion, pour des raisons de budget, de temps et de méthodologie. Cela nous laisse les indicateurs de l'erreur d'enquête totale et de ses composantes. Ces indicateurs n'ont qu'une valeur limitée pour les utilisateurs, qui ne savent que faire des taux de réponse, de la variance de réponse mesurée par répétition des interviews ou des taux de rejets au contrôle. Par contre, ils sont très utiles pour les producteurs des données d'enquête. Par exemple, des études par répétition des interviews permettent de déceler la falsification et les questions d'enquête pour lesquelles la réponse manque de cohérence. La majorité des utilisateurs apprécie la crédibilité du fournisseur de services et cette crédibilité tient en partie à la capacité de présenter des données exactes. Un autre aspect important de la crédibilité est la volonté qu'ont les fournisseurs d'évaluer leur propre qualité et de communiquer les résultats de ces

évaluations. Même si ces dernières révèlent des problèmes, il est préférable que ce soit le fournisseur qui les découvre plutôt que des entités externes. La plupart des utilisateurs ne souhaitent pas participer aux discussions au sujet des erreurs et des compromis entre les divers types d'erreurs, et ce, pour de bonnes raisons. Elles sont simplement trop techniques et obscures. Si nous admettons qu'un processus de bonne qualité est une condition préalable à un produit de bonne qualité, nous devrions améliorer progressivement les processus afin qu'ils s'approchent de la situation parfaite d'absence de biais. De cette façon, la variance d'une estimation devient une bonne approximation de l'erreur quadratique moyenne.

Malgré des discussions sans fin et une myriade de projets d'amélioration de la qualité des enquêtes, les pratiques n'ont guère changé (Lynn 2004 ; Pink, Borowik et Lee 2010 ; Groves 2011 ; Bohata 2011). Le manque de compétence au sein des organismes d'enquête est peut-être l'une des causes profondes de cette lenteur du changement. La recherche sur les enquêtes doit faire appel à de nombreuses théories et méthodologies, dont la statistique, la technologie de l'information, la gestion, la communication et les sciences du comportement. Ces dernières sont nécessaires pour déterminer les causes fondamentales des erreurs non dues à l'échantillonnage. Si l'on se contente de quantifier ces erreurs, aucune amélioration n'est possible. À l'heure actuelle, les programmes de formation insistent sur les erreurs d'échantillonnage, de non-réponse et de couverture, et sur l'estimation en présence de ces erreurs. D'autres processus et sources d'erreur, tels que la production de mesures et le traitement des données, ne se voient pas accorder autant d'importance. D'où une situation où les études sur l'erreur de mesure et sur l'erreur de traitement des données sont rares comparativement à celles sur disons, la non-réponse. Tant dans le camp des producteurs que dans celui des utilisateurs, la confusion est importante en ce qui concerne les concepts et les méthodes. Une autre cause de lenteur pourrait être la règle du consensus appliquée dans certains organismes lorsqu'il s'agit de prendre des décisions concernant les changements. Cette règle repose sur le compromis. L'avis de nombreuses parties prenantes est recueilli et une décision est habituellement prise en se basant sur le plus petit commun dénominateur, ce qui n'est jamais une bonne norme. En outre, arriver à ce compromis demande habituellement beaucoup de temps et de ressources. Cette approche est très éloignée du modèle planifier-faire-contrôler-agir.

La qualité des enquêtes n'est pas une entité absolue. Le mode uniformisé de communication de l'information sur la qualité en vigueur à l'heure actuelle ne convient pas, puisque chaque utilisateur définit l'adéquation à l'usage prévu. Les décisions concernant des dimensions de la

qualité telles que l'actualité, la comparabilité et l'accessibilité doivent être prises en collaboration avec les utilisateurs principaux, tandis que le fournisseur du service est responsable d'offrir la meilleure exactitude possible, compte tenu des diverses contraintes.

Les discussions sur la qualité des enquêtes et l'adoption de stratégies de gestion de la qualité ont-elles abouti à de meilleures données ? Nous ne le savons pas. La qualité des enquêtes n'a pas été évaluée selon un mode avant-après. La tendance est à l'accroissement de la normalisation et de la centralisation, ce qui devrait s'avérer rentable, mais quand il s'agit de la qualité des données, certains indicateurs pointent dans la mauvaise direction. Par exemple, dans de nombreux pays, les taux de non-réponse augmentent et les propriétés des erreurs dues à la collecte en mode mixte, à la traduction du matériel d'enquête et à d'autres caractéristiques de conception ne sont pas entièrement comprises ou varient d'une culture à l'autre. Il n'existe pas de formule de conception, ce qui entraîne des prises de décisions boiteuses concernant les compromis et des difficultés à décider de l'intensité avec laquelle les contrôles de qualité doivent être appliqués. La quête de pratiques exemplaires persiste dans les organismes d'enquête, mais leur mise en œuvre est difficile et éparpillée. Un rehaussement généralisé du niveau des compétences s'impose manifestement. Un programme de perfectionnement international structuré à l'intention des fournisseurs de services est nécessaire, de même qu'une collaboration internationale systématique en vue de déterminer les meilleurs moyens de concevoir et de mettre en œuvre les enquêtes. Nous devons mieux servir les utilisateurs en leur fournissant des données dont l'erreur est faible. Nous pouvons pour cela combiner plus judicieusement nos connaissances de la statistique et des phénomènes cognitifs avec les principes de gestion de la qualité. La note vraiment positive est l'attitude extraordinairement favorable à l'amélioration de la qualité dont témoignent les organismes statistiques partout dans le monde.

Bibliographie

- Aitken, A., Hörnigren, J., Jones, N., Lewis, D. et Zilhao, M. (2004). *Handbook on improving quality by analysis of process variables*. Office for National Statistics, Royaume-Uni.
- Anderson, R., Kasper, J. et Frankel, F. (1979). *Total Survey Error: Applications to Improve Health Surveys*. San Francisco : Jossey-Bass.
- Apted, L., Carruthers, P., Lee, G., Oehm, D. et Yu, F. (2011). *Industrialisation of statistical processes, methods and technologies*. Document présenté à la réunion de l'Institut International de Statistique, Dublin.

- Bailar, B., et Dalenius, T. (1969). Estimating the response variance components of the U.S. Bureau of the Census' Survey Model. *Sankhyā*, B, 341-360.
- Biemer, P. (2001). Commentaire sur Platek et Särndal. *Journal of Official Statistics*, 17(1), 25-32.
- Biemer, P. (2010). Overview of design issues: Total survey error. Dans *Handbook of Survey Research*, (Éds., P. Mardsen et J. Wright), Deuxième édition. Emerald Group Publishing Limited.
- Biemer, P., et Lyberg, L. (2003). *Introduction to Survey Quality*. New York : John Wiley & Sons, Inc.
- Biemer, P., et Lyberg, L. (2012). Short course on Total Survey Error. The Joint Program in Survey Methodology (JPSM), 16 au 17 avril, Washington, DC.
- Biemer, P., Trewin, D., Japac, L., Bergdahl, H. et Pettersson, Å. (2012). A tool for managing product quality. Document présenté à la conférence de Q2012, Athènes.
- Blyth, B. (2012). ISO 20252; Turning frameworks into best practice. Document présenté à la conférence de Q2012, Athènes.
- Bohata, M. (2011). Fit-for-purpose statistics for evidence based policy making. Note, Eurostat.
- Bowley, A.L. (1913). Working-class households in reading. *Journal of the Royal Statistical Society*, 76(7), 672-701.
- Box, G. (1990). Good quality costs less? How come? *Quality Engineering*, 3, 1, 85-90.
- Box, G., et Friends (2006). *Improving Almost Anything: Ideas and Essays*. New York : John Wiley & Sons, Inc.
- Brackstone, G. (1999). La gestion de la qualité des données dans un bureau de statistique. *Techniques d'enquête*, 25, 2, 159-171.
- Brackstone, G. (2001). Quelle est l'importance de l'exactitude ? *Recueil : Symposium 2001, La Qualité des Données d'un Organisme Statistique : Une Perspective Méthodologique*, Statistique Canada.
- Breyfogle, F. (2003). *Implementing Six Sigma*. Deuxième édition. New York : John Wiley & Sons, Inc.
- Brooks, C., et Bailar, B. (1978). An error profile: Employment as measured by the Current Population Survey. Document de travail 3, Office of Management and Budget, Washington, DC.
- Chakrabarty, R., et Torres, G. (1996). American Housing Survey: A Quality Profile. U.S. Department of Commerce, U.S. Bureau of the Census.
- Colledge, M., et March, M. (1993). Quality management: Development of a framework for a statistical agency. *Journal of Business and Economic Statistics*, 11, 157-165.
- Colledge, M., et March, M. (1997). Quality policies, standards, guidelines, and recommended practices. Dans *Survey Measurement and Process Quality*, (Éds., L. Lyberg, P. Biemer., M. Collins, E. De Leeuw, C. Dippo, N. Schwarz et D. Trewin), New York : John Wiley & Sons, Inc.
- Couper, M. (1998). Measuring Survey Quality in a CASIC Environment. Document présenté au Joint Statistical Meetings, American Statistical Association, Dallas, TX.
- Dalenius, T. (1967). Nonsampling Errors in Census and Sample Surveys. Rapport N° 5 du projet de recherche Errors in Surveys. Stockholm University.
- Dalenius, T.E. (1968). Official statistics and their uses. *Revue de l'Institut International de Statistique*, 26(2), 121-140.
- Dalenius, T. (1969). Designing descriptive sample surveys. Dans *New Developments in Survey Sampling*, (Éds., N.L. Johnson et H. Smith), New York : John Wiley & Sons, Inc.
- Dalenius, T. (1985a). *Elements of Survey Sampling*. Swedish Agency for Research Cooperation with Developing Countries. Stockholm, Suède.
- Dalenius, T. (1985b). Relevant official statistics. *Journal of Official Statistics*, 1(1), 21-33.
- Deming, E. (1944). On errors in surveys. *American Sociological Review*, 9, 359-369.
- Deming, E. (1950). *Some Theory of Sampling*. New York : John Wiley & Sons, Inc.
- Deming, E. (1986). *Out of the Crisis*. MIT.
- Deming, W.E., et Geoffrey, L. (1941). On sample inspection in the processing of census returns. *Journal of the American Statistical Association*, 36, 215, 351-360.
- De Vries, W. (1999). Are we measuring up...? Questions on the performance of national systems. *Revue Internationale de Statistique*, 67, 1, 63-77.
- Dillman, D. (1996). Why innovation is difficult in government surveys (avec discussions). *Journal of Official Statistics*, 12, 2, 113-198.
- Doherty, K. (2010). How business architecture renewal is changing IT at Statistics Canada. Document présenté au Meeting on the Management of Statistical Information Systems. Daejeon, Corée du Sud, 26 au 29 avril.
- Doyle, P., et Clark, C. (2001). Quality profiles and data users. Document présenté à l'International Conference on Quality in Official Statistics (Q), Stockholm.
- Drucker, P. (1985). *Management*. Harper Colophone.
- Ecochard, P., Hahn, M. et Junker, C. (2008). User satisfaction surveys in Eurostat and in the European Statistical System. Document présenté à la conférence Q, Rome, Italie.
- Edwards, W., Lindman, H. et Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193-242.
- Eltinge, J. (2011). Aggregate and systemic components of risk in total survey error models. Document présenté au ITSEW 2011, Québec, Canada.

- Ericson, W. (1969). Subjective Bayesian models in sampling finite populations. *Journal of the Royal Statistical Society, Séries B*, 195-233.
- European Foundation for Quality Management (1999). *The EFQM Excellence Model*. Van Haren.
- Eurostat (2009a). ESS Standard for Quality Reports. Eurostat.
- Eurostat (2009b). ESS handbook for Quality Reports. Eurostat.
- Eurostat (2011a). European statistics Code of Practice. Eurostat.
- Eurostat (2011b). Quality assurance framework (QAF). Eurostat.
- Fellegi, I. (1964). Response variance and its estimation. *Journal of the American Statistical Association*, 59, 1016-1041.
- Fellegi, I. (1996). Characteristics of an effective statistical system. *Revue Internationale de Statistique*, 64, 2, 165-197.
- Felme, S., Lyberg, L. et Olsson, L. (1976). *Kvalitetsskydd av data. (Protecting Data Quality.) Liber* (en suédois).
- Fienberg, S., et Tanur, J. (1996). Reconsidering the fundamental contributions of Fisher and Neyman on experimentation and sampling. *Revue Internationale de Statistique*, 64, 237-253.
- Fisher, R. (1935). *The Design of Experiments*. New York : Hafner.
- Frankel, M., et King, B. (1996). A conversation with Leslie Kish. *Statistical Science*, 11, 1, 65-87.
- Gleaton, E. (2011). Centralizing LAN services. Note, National Agricultural Statistics Service, U.S. Department of Agriculture.
- Groves, R. (1989). *Survey Errors and Survey Costs*. New York : John Wiley & Sons, Inc.
- Groves, R. (2011). The structure and activities of the U.S. Federal Statistical System: History and recurrent challenges. *The Annals of the American Academy of Political and Social Science*, 631, 163, Sage.
- Groves, R., Biemer, P., Lyberg, L., Massey, J., Nicholls, W. et Waksberg, J. (Éds.) (1988). *Telephone Survey Methodology*. New York : John Wiley & Sons, Inc.
- Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. et Tourangeau, R. (2009). *Survey Methodology*, Deuxième édition. New York : John Wiley & Sons, Inc.
- Groves, R., et Heeringa, S. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society, A*, 169, 439-457.
- Groves, R., et Lyberg, L. (2010). Total survey error: Past, present and future. *Public Opinion Quarterly*, 74, 5, 849-879.
- Hansen, M., et Hurwitz, W. (1946). The problem of nonresponse in sample surveys. *Journal of the American Statistical Association*, 517-529.
- Hansen, M., Hurwitz, W. et Bershad, M. (1961). Measurement errors in censuses and surveys. *Bulletin of the International Statistical Institute*, 32^e Session, 38, Partie 2, 359-374.
- Hansen, M., Hurwitz, W. et Madow, W. (1953). *Sample Survey Methods and Theory*. Volumes I et II. New York : John Wiley & Sons, Inc.
- Hansen, M., Hurwitz, W., Marks, E. et Mauldin, P. (1951). Response errors in surveys. *Journal of the American Statistical Association*, 46, 147-190.
- Hansen, M., Hurwitz, W. et Pritzker, L. (1964). The estimation and interpretation of gross differences and simple response variance. Dans *Contributions to Statistics*, (Éd., C. Rao). Oxford : Pergamon Press, 111-136.
- Hansen, M., Hurwitz, W. et Pritzker, L. (1967). Standardization of procedures for the evaluation of data: Measurement errors and statistical standards in the Bureau of the Census. Document présenté à la réunion de l'Institut International de Statistique, Sydney.
- Hansen, M., et Steinberg, J. (1956). Control of errors in surveys. *Biometrics*, 462-474.
- Hansen, M., et Voigt, R. (1967). Program guidance through the evaluation of uses of official Statistics in the United States Bureau of the Census. Document présenté à la réunion de l'Institut International de Statistique, Sydney.
- Holt, T., et Jones, T. (1998). Quality work and conflicting policy objectives. *Proceedings of the 84th DGINS Conference*, 28 au 29 mai, Stockholm, Suède. Eurostat.
- Jabine, T., King, K. et Petroni, R. (1990). Survey of Income and Program Participation (SIPP): Quality Profile. U.S. Department of Commerce, U.S. Bureau of the Census.
- Joiner, B. (1994). *Generation Management*. McGraw-Hill.
- Julien, C., et Born, A. (2006). Quality management assessment at Statistics Canada. *Proceedings of the Q Conference*, Cardiff, Royaume-Uni.
- Julien, C., et Royce, D. (2007). Quality review of key indicators at Statistics Canada. *Proceedings of the Third International Conference on Establishment Surveys (ICES-III)*, 1113-1120.
- Juran, J.M. (1988). *Juran on Planning for Quality*. New York : Free Press.
- Juran, J.M. (1995). *A History of Managing for Quality*. ASQC Quality Press.
- Juran, J., et Gryna, F. (Éds.) (1988). *Juran's Quality Control Handbook*, 4^e édition. McGraw-Hill.
- Kalton, G. (2001). Quelle est l'importance de l'exactitude ? *Recueil : Symposium 2001, La Qualité des Données d'un Organisme Statistique : Une Perspective Méthodologique*, Statistique Canada.
- Kalton, G., Winglee, M. et Jabine, T. (1998). *SIPP Quality Profile*. U.S. Bureau of the Census, 3^e édition.
- Kalton, G., Winglee, M., Krawchuk, S. et Levine, D. (2000). *Quality Profile for SASS Rounds 1-3: 1987-1995*. Washington, DC : U.S. Department of Education.
- Kasprzyk, D., et Kalton, G. (2001). Quality profiles in U.S. Statistical Agencies. *Proceedings of the International Conference on Quality in Official Statistics*, Stockholm 14 au 15 mai 2001, CD-ROM.

- Kennickell, A., Mulrow, E. et Scheuren, F. (2009). Paradata or process modeling for inference. Document présenté à la Conférence on Modernization of Statistics Production, Stockholm, Suède.
- Kiear, A.N. (1897). The representative method of statistical surveys. *Kristiania Videnskaps-selskabets Skrifter: Historik-filosofiske Klasse*, (en norvégien), 4, 37-56.
- Kish, L. (1965). *Survey Sampling*. New York : John Wiley & Sons, Inc.
- Kish, L. (1995). *The Hundred Years' Wars of Survey Sampling*. Centennial Representative Sampling, Rome.
- Kotz, S. (2005). Reflections on early history of official statistics and a modest proposal for global coordination. *Journal of Official Statistics*, 21, 2, 139-144.
- Kreuter, F., Couper, M. et Lyberg, L. (2010). The use of paradata to monitor and manage survey data collection. *Proceedings of the Survey Research Methods Section*, American Statistical Association.
- Lyberg, L. (1981). *Control of the Coding Operation in Statistical Investigations: Some Contributions*. Thèse de doctorat, Stockholm University.
- Lyberg, L. (2002). Training of survey statisticians in government agencies-A review. Communication sollicitée présentée à la réunion des Joint Statistical Meetings, American Statistical Association, New York.
- Lyberg, L., Bergdahl, M., Blanc, M., Booleman, M., Grünewald, W., Haworth, M., Japac, L., Jones, L., Körner, T., Linden, H., Lundholm, G., Madaleno, M., Radermacher, W., Signore, M., Zilhao, M.J., Tzougas, I. et van Brakel, R. (2001). Summary report from the Leadership Group (LEG) on Quality. Eurostat.
- Lyberg, L., et Couper, M. (2005). The use of paradata in survey research. Communication sollicitée présentée à la réunion de l'Institut International de Statistique, Sydney.
- Lynn, P. (Éd.) (2003). *Quality Profile: British Household Panel Survey: Waves 1 to 10: 1991-2000*. Colchester : Institute for Social and Economic Research.
- Lynn, P. (2004). Editorial: Measuring and communicating survey quality. *Journal of the Royal Statistical Society, Séries A*, 167.
- Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, 109, 325-378.
- Mirotschie, M. (1993). Data quality: A quest for standard indicators. *Proceedings of the International Conference on Establishment Surveys*, American Statistical Association, 729-734.
- Moeller, R. (2005). *Brink's Modern Internal Auditing*. Sixième édition. New York : John Wiley & Sons, Inc.
- Morganstein, D., et Marker, D. (1997). Continuous quality improvement in statistical agencies. Dans *Survey Measurement and Process Quality*, (Éds., L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz et D. Trewin), New York : John Wiley & Sons, Inc., 475-500.
- Mudryk, W., Burgess, M.J. et Xiao, P. (1996). Quality control of CATI operations in Statistics Canada, Note, Statistique Canada.
- Muscio, B. (1917). The influence of the form of a question. *The British Journal of Psychology*, 8, 351-389.
- Nations Unies (2010). *Post Enumeration Surveys: Operational Guidelines*. Department of Economic and Social Affairs, Statistics Division.
- Neter, J., et Waksberg, J. (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association*, 59, 305, 18-55.
- Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-606.
- Neyman, J. (1938). *Lectures and Conferences on Mathematical Statistics and Probability*. U.S. Department of Agriculture, Washington, DC.
- OCDE (2011). Quality dimensions, core values for OECD statistics and procedures for planning and evaluating statistical activities. OCDE.
- O'Muirheartaigh, C. (1997). Measurement errors in surveys: A historical perspective. Dans *Survey Measurement and Process Quality*, (Éds., L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz et D. Trewin), New York : John Wiley & Sons, Inc., 1-25.
- Organisation internationale de normalisation (2006). Market, Opinion and Social Research. ISO Standard N° 20252.
- Phipps, P., et Fricker, S. (2011). Quality measures. Note, Office of Survey Methods Research, U.S. Bureau of Labor Statistics.
- Pink, B., Borowik, J. et Lee, G. (2010). The case for an international statistical innovation program-Transforming national and international statistics systems. Document présenté au Collaboration Leaders Workshop, 19 au 23 avril, Sydney, Australie.
- Platek, R., et Särndal, C.-E. (2001). Can a statistician deliver? *Journal of Official Statistics*, 17, 1, 1-20 et la discussion, 21-27.
- Reedman, L., et Julien, C. (2010). Current and future applications of the generic statistical business process model at Statistics Canada. Document présenté au Q Conference, Helsinki.
- Rosén, B., et Elvers, E. (1999). Quality concept for official statistics. *Encyclopedia of Statistical Sciences*, New York : John Wiley & Sons, Inc., mise à jour, Volume 3, 621-629.
- Scheuren, F. (2001). Quelle est l'importance de l'exactitude ? *Recueil : Symposium 2001, La Qualité des Données d'un Organisme Statistique : Une Perspective Méthodologique*, Statistique Canada.
- Schilling, E., et Neubauer, D. (2009). *Acceptance Sampling in Quality Control*, 2^e éd. Chapman and Hall/CRC.

- Scholtes, P., Joiner, B. et Streibel, B. (1996). *The Team Handbook*. Joiner Associates Inc.
- Shewhart, W.A. (1939). *Statistical Methods from the Viewpoint of Quality Control*. U.S. Department of Agriculture, Washington, DC, États-Unis.
- Smith, T. (2011). Report on the International Workshop on using multi-level data from sample frames, auxiliary databases, paradata and related sources to detect and adjust for nonresponse bias in surveys. NORC/University of Chicago.
- Spencer, B. (1985). Optimal data quality. *Journal of the American Statistical Association*, 80, 564-573.
- Statistique Canada (2002). Le cadre d'assurance de la qualité de Statistique Canada, N° au catalogue 12-586-XIE, Ottawa.
- Statistique Canada (2009). Statistique Canada : Lignes directrices concernant la qualité, cinquième édition, Ottawa.
- Statistics Netherlands (1997). A self assessment of the Department of Statistical Methods. Document de recherche N° 9747, Statistics Netherlands.
- Stephan, F.F. (1948). History of the uses of modern sampling procedures. *Journal of the American Statistical Association*, 43, 12-39.
- Trewin, D. (2001). Importance d'une culture de la qualité. *Recueil : Symposium 2001, La Qualité des Données d'un Organisme Statistique : Une Perspective Méthodologique*, Statistique Canada.
- U.S. Bureau of the Census (1974). *Standards for Discussion and Presentation of Errors in Data*. U.S. Department of Commerce, Bureau of the Census.
- U.S. Federal Committee on Statistical Methodology (2001). *Measuring and Reporting Sources of Errors in Surveys*, Statistical Policy, document de travail 31, Washington, DC : U.S. Office of Management and Budget.
- U.S. Office of Management and Budget (2002). Guidelines for ensuring, and maximizing the quality, objectivity, utility, and integrity of information disseminated by Federal agencies. Federal register, 67, 36, 22 février.
- U.S. Office of Management and Budget (2006a). *Standards and Guidelines for Statistical Surveys*. U.S. Office for Management and Budget.
- U.S. Office of Management and Budget (2006b). Questions and answers when designing surveys for information collection. U.S. Office for management and Budget.
- Waksberg, J. (1998). The Hansen era: Statistical research and its implementation at the Census Bureau, 1940-1970. *Journal of Official Statistics*, 14, 2, 119-137.
- Weisberg, H. (2005). *The Total Survey Error Approach*. The University of Chicago Press.
- Weisman, E., Balyozov, Z. et Venter, L. (2010). IMF's data quality assessment framework. Document présenté à la Conference on Data Quality for International Organizations, Helsinki, 6 au 7 mai.
- West, B., et Olson, K. (2010). How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly*, 74, 5, 1004-1026.
- Willimack, D., Nichols, E. et Sudman, S. (2002). Understanding unit and item nonresponse in business surveys. Dans *Survey Nonresponse*, (Éds., R. Groves, D. Dillman, J. Eltinge et R. Little), 213-228.
- Zarkovich, S. (1966). *Quality of Statistical Data*. Food and Agricultural Organization of the United Nations : Rome, Italie.