

## BANQUE DE GESTION D'APPELS DE SONDAGE TÉLÉPHONIQUE – UTILISATION ET VALIDATION

Serge Chevalier<sup>1</sup>

### RÉSUMÉ

La banque de gestion d'appels contient le résultat de chacun des appels de tous les numéros de téléphone échantillonnés dans un sondage téléphonique. La base inclut une série d'informations dont celles relatives au résultat de l'appel circonscrit en une trentaine de catégories (entrevue complétée, refus, rendez-vous, ligne occupée, etc.). Les données disponibles dans une telle banque permettent d'établir le taux de productivité de l'échantillon (associé aux coûts du projet et au nombre de répondants au sondage), le taux de réponse au sondage (un indice de la qualité des résultats obtenues) et les pondérations. Il sera fait état de l'importance théorique de la présente démarche. Des éléments de validation nécessaires à l'utilisation optimale de ces informations – la validation sera développée autour de concepts théoriques. Des résultats comparatifs, avant et après validation, montreront des écarts en regard du taux de productivité et du taux de réponse. Des solutions préliminaires seront proposées.

MOTS CLÉS : Méthodologie de sondage, banque de gestion d'appel, validation, taux de réponse, taux de productivité.

### ABSTRACT

Call sheets data contains the result of every phone call made to achieve a survey. The data includes many types of information; for our purpose only the informations pertaining the call results are studied here (completed questionnaire, refusal, no answer, etc.). The available data permit the computation of the productivity rate (associated with the cost of the survey and the number of respondents), the response rate (a measure of the quality of the gathered data), and the weighting of the raw data. Theoretical elements will be presented as will the validation scheme of the data. Results indicate discrepancies between the productivity rate and the response rate computed before and after validation. Elements of solution will be discussed.

KEY WORDS: Survey methodology, Call sheets data, Validation, Hit rate, Response rate.

### 1. INTRODUCTION

À un moment ou à un autre de la démarche d'un sondage téléphonique on en vient à échantillonner, d'une manière ou d'une autre, un certain nombre de numéros de téléphone. Chaque numéro doit être composé afin de tenter de rejoindre une personne ou un ménage. La banque de gestion d'appels (ou registre des appels) constituée alors sert à faire le suivi de chacun des appels. Il s'agit d'une procédure dont on ne peut faire l'économie; toutes les enquêtes sérieuses procèdent de cette manière. Lorsque le terrain est complété, la banque de gestion d'appels sert

directement au calcul du taux de productivité ainsi qu'au taux de réponse et contribue au calcul des pondérations. Malgré l'importance de cette procédure très peu d'attention y a été portée tant sur le plan théorique qu'en termes pratiques. Groves (1989) propose une nomenclature du résultat des appels et indique comment calculer un taux de réponse, il va même jusqu'à mentionner que le classement final d'un appel est sujet à une erreur de mesure. Lavrakas (1993) fournit des indications sur le contenu minimal d'un registre des appels, il offre, à l'instar de Groves, une nomenclature du résultat des appels et une méthode similaire de calcul du taux de réponse. Notre

<sup>1</sup> Serge Chevalier, Régie régionale de la santé et des services sociaux de Montréal-Centre, Direction de la santé publique et Institut national de santé publique du Québec, 1301, rue Sherbrooke Est, Montréal, (Qc) H2L 1M3; courrier électronique : [scheval2@santepub-mtl.qc.ca](mailto:scheval2@santepub-mtl.qc.ca)

recherche vise un triple objectif : d'abord, développer une méthodologie de validation de la banque de gestion d'appel, ensuite, mesurer l'ampleur de l'erreur de mesure qu'on y retrouve et, finalement, proposer une typologie des erreurs retrouvées, de leurs impacts et d'identifier certaines pistes de solution visant à en éviter le plus possible.

## 2. THÉORIE

La banque de gestion d'appel fournit l'ensemble des informations nécessaires au calcul du taux de productivité, du taux de réponse ainsi que des renseignements indispensables au calcul des pondérations. À chacun de ces concepts se rattachent des erreurs différentes aux conséquences distinctes.

Le taux de productivité se définit comme la proportion des numéros de téléphone échantillonné correspondant à la population visée; par exemple ce peut être les ménages privés ou encore les ménages privés dans lesquels on retrouve un adolescent âgés de 13 à 17 ans. Le taux de productivité sert à établir la taille de l'échantillon. L'utilisation d'un taux de productivité trop élevé, par rapport à la réalité, conduit à surestimer le nombre d'éléments de la population visée dans l'échantillon et, par extension, du nombre de répondants. Il en découle que la précision des estimateurs peut être affectée ou même que certaines analyses prévues ou nécessaires puissent être impossibles à produire.

Le taux de réponse représente un des indices les plus connus et utilisés dans le monde des sondages. Il s'agit d'un indice global de qualité. Plus le taux de réponse s'approche de 100 %, moins la possibilité d'être en présence d'un échantillon biaisé existe. Plus le taux de réponse est faible, moins le sondage est crédible et plus la démonstration de la représentativité de l'échantillon repose sur les épaules des chercheurs. De manière indirecte, le taux de réponse influence le calcul des pondérations – en fait, c'est la quantité et la nature des non-répondants qui affectent les pondérations, pas le taux de réponse en soi. Le taux de réponse est obtenu en divisant le nombre de personnes ou de ménages répondants par le nombre de personnes ou de ménages appartenant à la population visée (i.e. la somme des répondants, des refus ainsi que des personnes ou ménages qui n'ont pas été rejoints).

Au chapitre des pondérations, nous savons qu'elles sont affectées par la quantité et la nature de la non-réponse par rapport à la réponse. Les pondérations affectent, marginalement ou de manière importante selon le sondage, les estimateurs; le cas échéant, les

relations entre les variables peuvent aussi être perturbées.

## 3. MÉTHODOLOGIE

La méthodologie de validation de la banque de gestion d'appel que nous avons développée a été définie théoriquement. Nous avons commencé par identifier tous les moyens externes de validation de chacun des codes. Par exemple, vérifier si à chacune des occurrences de code d'entrevue complétée correspond une entrevue complétée ou encore vérifier si des codes de résultats inconnus sont utilisés ou encore si le code final, celui du dernier appel réalisé, est plausible. Le second volet consiste à établir une liste exhaustive de toutes les séquences inadmissibles de code (par exemple, un code de premier refus de ménage qui suit un code de deuxième refus de ménage ou encore un code «pas de réponse» au lieu de «pas de réponse après un refus»); au total, une centaine de séquences inadmissibles a ainsi été identifiées. Le modèle théorique des erreurs a ensuite été appliquée à une enquête en cours de réalisation, l'Enquête de santé mentale auprès des Montréalais (Fournier et al., sous presse).

Il s'agit d'une enquête téléphonique réalisée en 1999, sur 16 semaines et pour laquelle nous avons retenu les services d'une firme spécialisée en vue de la gestion du terrain et de l'administration, assistée par ordinateur, du questionnaire. Nous avons utilisé une méthodologie de génération aléatoire de numéro de téléphone et sélectionné, au hasard, une personne par ménage parmi celles éligibles. Nous visons la population montréalaise âgée de 18 ans et plus; un questionnaire long a été administré à près de 95 % des répondants; pour les cas où la personne sélectionnée s'avérait inapte à compléter l'entrevue à cause de problème de santé (physique ou mentale) ou de langue nous avons demandé, là où cela était possible, à un colocataire, de la même famille ou non, connaissant bien la personne sélectionnée de répondre à un questionnaire beaucoup plus court portant sur la situation de la personne sélectionnée. Malgré que les informations recueillies chez un tiers soient moins fiables que celles obtenues de la personne concernée, il valait mieux, pour nous, obtenir certains renseignements plutôt qu'aucun. Trois sous-groupes de la population nous semblaient incontournables à enquêter : 1<sup>o</sup> les personnes ayant un problème de santé mentale suffisamment grave pour les empêcher de répondre, 2<sup>o</sup> les personnes atteintes de problèmes de santé physique assez graves pour les empêcher de répondre, sachant que de tels problèmes de santé physique sont souvent associés à des problèmes de

santé mentale et 3° les personnes qui ne peuvent s'exprimer ni en français ni en anglais – à Montréal, les personnes de langue maternelle autre que le français et l'anglais représentent autour du quart de la population et il s'agit d'un sous-groupe jugé prioritaire par les autorités sanitaires. Nous avons obtenu les réponses de 4 704 personnes à partir d'un échantillon de 12 102 numéros de téléphone échantillonnés qui ont nécessité 112 310 appels réalisés par 48 intervieweurs.

Nous sommes convenus avec la firme d'une liste de plus d'une trentaine de codes de résultat d'appel. Ces codes peuvent être regroupés en cinq catégories principales : 1° entrevue complétée, 2° ménage ou personne qui n'a pas été rejoint, 3° refus du ménage ou de la personne sélectionnée, 4° numéro de téléphone invalide (non-résidentiel, télécopieur, pas de service, téléphone cellulaire, téléavertisseur) et 5° numéro inadmissible (dans le cadre de notre enquête, seuls les numéros correspondant à un ménage où ne réside aucune personne âgée de 18 ans et plus sont inadmissibles).

#### 4. RÉSULTATS

La liste des codes inadmissibles, des codes erronés et des séquences inadmissibles de code dépend des codes de résultat retenus ou adoptés – ces codes peuvent varier d'une enquête à une autre. Nous ne présenterons pas ici la liste que nous avons produite. Une telle liste est relativement facile à établir. Un exemple suffira à montrer le type de démarche nécessaire à l'établissement de la liste. Supposons un code final «pas de réponse», un tel code appartient à la famille des «pas rejoint»; quels sont les codes qui peuvent apparaître avant un tel code final ou, inversement, qui ne devraient pas apparaître avant ce code final. En consultant la liste des codes possibles nous pouvons déterminer que, si tous les codes précédents sont aussi des «pas de réponse», il s'agit d'un code final permis, de même si des codes «ligne occupée» s'intercalent parmi des codes «pas de réponse»; à l'opposé, si un code indiquant un refus apparaît avant un code final de «pas de réponse», nous sommes en présence d'une erreur – il s'agit d'une erreur de classification si le code «refus» est adéquat. En procédant ainsi pour tous les codes possibles, nous obtenons une classification à trois modalités : 1° absence d'erreur détectable, 2° erreur détectée 3° possibilité d'erreur, vérification supplémentaire à effectuer.

Après avoir appliqué la grille des erreurs possibles au fichier de gestion des appels de l'enquête, nous

observons une proportion de 15,4 % d'erreurs de classification, c'est-à-dire une erreur détectable pour 1 869 des 12 102 numéros de téléphone sélectionnés. Toutes les erreurs ne revêtent pas la même importance. Par exemple, un code final de refus du ménage pourrait apparaître au lieu d'un code de refus de la personne sélectionnée; auquel cas le code final appartient à la grande catégorie des refus et l'impact d'une telle erreur est nulle quant au calcul des taux de productivité et de réponse et probablement aussi au chapitre du calcul des pondérations. Une telle erreur n'a d'importance que sur le plan des apprentissages et des adaptations en vue de futures enquêtes; les mesures à implanter pour réduire la non-réponse des ménages ne sont pas les mêmes que celles visant l'atténuation de la non-réponse des personnes sélectionnées. Les erreurs les plus conséquentes se sont produites pour 7,2 % (873) des numéros échantillonnés (tableau 1). Deux grands types d'erreurs s'avèrent plus prévalants. La classification finale «pas rejoint» affiche la plus grande proportion relative d'erreur alors que seulement 152 des 697 numéros se sont avérés bien classés; autrement 319 des 4 848 catégorisations initiales «invalide» ont dû être révisées. L'impact de ces erreurs peut aussi être très variable ou même impossible à évaluer.

Le taux de productivité qui avait été annoncé s'élevait à 65 %; le taux calculé après avoir corrigé les erreurs identifiés atteint 58,6 %. En l'occurrence l'erreur relative approche les 10 % (tableau 2).

Au chapitre des taux de réponse nous pouvons en calculer trois : celui du questionnaire long, celui du questionnaire administré au tiers et un taux global pour les deux instruments. Tous les taux établis après correction s'avèrent plus faible que ceux calculés à partir des résultats initiaux. L'ampleur des différences relatives varie d'environ 2 % à plus de 40 %. Le taux de réponse global après correction s'établit à 2 % de moins que celui fourni par la firme.

Trois catégories de codes finaux revêtent une importance déterminante dans le calcul des pondérations : le nombre de refus, le nombre de personne ou de ménage qui n'a pas été rejoint et le nombre de questionnaire complété par des efforts de récupération (après au moins un refus). Le calcul ajusté montre que les refus et les questionnaires complétés en récupération ont été sous-évalués alors que les «pas rejoints» sont surévalués. Les erreurs relatives varient de 17 % à 160 %.

**Tableau 1. Classification initiale fournit dans la banque de gestion d'appel et classification révisée pour chacun des numéros de téléphone échantillonnés**

Classification de la firme	Classification révisée					Total
	Complété	Pas rejoint	Refus	Invalide	Inadmissible	
Complété	4703		3			4706
Pas rejoint	1	152	66	478		697
Refus			1808	6		1814
Invalide		75	244	4529		4848
Inadmissible					37	37
<b>Total</b>	<b>4704</b>	<b>227</b>	<b>2121</b>	<b>5013</b>	<b>37</b>	<b>12102</b>

L'examen de l'ensemble des erreurs de classification a permis d'en détecter une quinzaine de type qui peuvent être regroupés en cinq catégories (figure 1).

La communication entre les chercheurs ou promoteurs d'une enquête et la firme s'avère un processus où peuvent être évité ou généré un certain nombre d'erreurs. Les chercheurs peuvent omettre de transmettre certaines informations ou décisions méthodologiques qui ont des incidences sur la collecte des données. Un exemple tiré de notre enquête est en rapport avec les numéros de téléphone qui correspondent à des télécopieurs. Nous pensions qu'après l'obtention de cinq codes de télécopieur (s'il n'y avait aucun autre code hormis des «lignes occupées») nous étions en présence d'une ligne non-

résidentielle, que le code final devait en être un d'inéligible et que les efforts pour obtenir une réponse à ce numéro devaient cesser. En cette matière, il n'existe pas de référence absolu, certains se contentent de quatre occurrences, d'autres proposent d'autres exigences. Or, nous n'avons pas fourni nos critères de décisions à la firme, ce qui aurait été dans notre intérêt commun. Inversement, toute une série de décisions rapides doivent être prises en cours de terrain. Les firmes hésitent à contacter les chercheurs pour prendre une décision ou valider une décision qui n'aura pu être reportée. Une telle manière de procéder peut aussi résulter en des conséquences fâcheuses, qui vont de la perte de temps à l'obtention de données de qualité insuffisante.

**Tableau 2. Impact des erreurs de classification sur le taux de productivité, les taux de réponse et certains résultats dont dépendent le calcul des pondérations**

	Annoncé	Calcul ajusté	Différence absolue	Différence relative
<b>Taux de productivité</b>	65,0 %	58,6 %	6,4 %	9,8 %
<b>Taux de réponse</b>				
Global	68,8 %	66,7 %	2,1 %	3,1 %
Questionnaire long	68,3 %	67,2 %	1,1 %	1,9 %
Questionnaire des tiers	97,2 %	57,7 %	39,5 %	40,6 %
<b>Pondération</b>				
Refus	1814	2121	-307	-16,9 %
Pas rejoint	697	227	470	67,4 %
Complété par récupération	449	1166	-717	-159,7 %

**Figure 1. Typologie des sources d'erreurs dans la banque de gestion d'appels**

- Communication entre chercheurs et firme
- Intervieweurs
  - formation générale
  - formation spécifique
  - erreurs humaines
- Logiciel (gestion du terrain & des entrevues)
  - limites du logiciel
  - programmation du logiciel
- Procédures de la firme
  - procédures de validation
  - régie interne
- Estimation du taux de productivité

Une série d'erreurs sont imputables aux intervieweurs eux-mêmes. Il n'existe pas de formation générale de base qui permette d'accéder à l'emploi d'intervieweur. Une grande partie des apprentissages se font sur le terrain, à l'usage et les firmes de sondage ont aussi un roulement important de personnel. Or, une formation insuffisante ou défailante peut entraîner des erreurs nombreuses, voire même, dans certains cas, systématiques. Pour diverses raisons, économiques aussi bien que pratiques, la formation spécifique relative à une enquête particulière peut aussi s'avérer inadéquate. Ainsi, les intervieweurs peuvent avoir à être formé en cours de terrain, la formation sera alors, le plus souvent, écourtée. Finalement, les intervieweurs ne sont pas à l'abri d'une mauvaise compréhension des directives; ils peuvent aussi, tout bonnement, à l'occasion, commettre des fautes de frappe à la saisie de l'information. Dans l'enquête qui nous occupe, l'erreur la plus fréquente et la plus importante a été commise par les intervieweurs. Les intervieweurs ont presque systématiquement omis d'utiliser ou ne connaissaient les deux codes «pas de réponse après refus» et «ligne occupée après refus». De telle sorte, qu'un grand nombre de codes faux auraient dû être des refus et se sont avérés des numéros qui n'ont pas été rejoints.

La troisième source d'erreur est le logiciel utilisé pour la gestion du terrain ainsi que pour l'administration des questionnaires. Les meilleurs logiciels sur le marché sont, à notre sens, les plus flexibles; ceux qui permettent de tenir compte du plus grand nombre de paramètres et qui, aussi, facilitent la programmation de modules ou de sous-routines correspondant à des besoins particuliers d'une ou de plusieurs enquêtes. Ce genre de logiciel a aussi des limites et est souvent

lourd d'utilisation. Dans le cas de notre enquête, il demeure encore difficile, après vérification, dans quelle mesure le logiciel utilisé permet d'implanter des tests efficaces pour valider la banque de gestion d'appel au fur et à mesure que les données sont colligées.

La quatrième source d'erreur réside dans les procédures de la firme. La firme implante un processus de vérification et de validation de la qualité des données. Elle peut ou non inclure des éléments de validation portant sur la banque de gestion d'appel. Autrement, les règles de régie interne peuvent aussi avoir un impact sur la qualité des données du registre des appels. Toujours dans le cas qui nous occupe, la firme avait avisé ses intervieweurs, bien avant le terrain de notre enquête, que les codes relatifs à la récupération ne seraient plus utilisés. Les intervieweurs embauchés après cette décision n'ont pas été informés de l'existence de ces codes. Ces codes n'ont cependant pas été désactivés. En d'autres termes, certains intervieweurs continuaient de les utiliser alors que d'autres en ignoraient tout simplement l'existence. Or, ces codes nous importent, pour le calcul des pondérations (selon la méthode que nous avons retenue). Ici, tout autant l'interface entre les chercheurs et la firme que les procédures sont en jeu.

La cinquième et dernière source d'erreur est en rapport à l'estimation du taux de productivité. La firme de sondage sous-contracte généralement cet aspect du sondage à une firme spécialisée en la matière et a relativement peu de contrôle en cette matière. Le taux de productivité connu pour l'ensemble du Québec gravite autour de 65 %, d'après le fournisseur, la

région de Montréal aurait un même taux. Ce qui est contestable. En bout de piste, nous avons obtenu un taux largement inférieur qui nous a empêché d'obtenir le nombre visé de répondants pour lequel nous étions engagés contractuellement.

## 5. DISCUSSION

Nous avons procédé à la validation du registre des appels pour une enquête téléphonique d'envergure. Pour ce faire nous avons dû développer des critères et des méthodes de validation. La stratégie retenue est fonction du nombre de codes de résultat d'appel utilisé et du degré de précision requis. Nous ne proposons pas de recette à cet égard car nous pensons qu'il faille développer des approches spécifiques à chaque enquête.

Nous avons dénombré 15,4 % d'erreurs de classification dans les 12 102 numéros échantillonnés; 7,2 % d'erreurs conséquentes. Il s'agit d'une proportion qui nous semble élevée – bien que nous ne disposions d'aucun point de comparaison. Ces erreurs affectent le calcul du taux de réponse, du taux de productivité et celui des pondérations. Il ne peut être exclu à ce moment, qu'une portion des erreurs dénombrées soient dues à la complexité du devis utilisé. Il se peut aussi que le taux soit anormalement élevé pour des raisons fortuites ou tout simplement aléatoires hors du contrôle des chercheurs et de la firme. Il faut cependant souligner que nous n'avons identifié que des erreurs qui, par définition, sont détectables. Or, il existe sûrement des erreurs que nous n'avons pu retracer. Ces dernières sont évidemment d'une quantité inconnue et ont un impact indéterminé.

Cinq types d'erreurs sont présentés. Les erreurs dues à une communication déficiente entre les chercheurs et la firme sont corrigibles dans une certaine mesure. Il s'agit, pour chacune des deux parties de bien faire connaître ses besoins d'information et de bien délimiter, *a priori*, le champ de compétence attribué à l'un et à l'autre. Eu égard aux intervieweurs, les chercheurs n'ont un accès direct à eux que pendant la

formation spécifique à leur enquête, le reste demeure sous le contrôle de la firme. Il est néanmoins possible de contrôler la qualité de leur travail par des mesures de validation informatisée au fil de la saisie des données et *a posteriori*. Le problème des logiciels de gestion de terrain et de gestion des entrevues est plus épineux. Là où le marché de ces logiciels est particulièrement concentré, plusieurs sondages peuvent être entachés d'erreurs similaires mais le nombre de correctifs à appliquer est moindre. Là où le marché est plus fragmenté, les erreurs seraient peut-être moins nombreuses en quantité, mais plus diversifiées et les correctifs plus difficiles à implanter. Finalement, les estimations de taux de productivité émanent d'un nombre fort limité de fournisseurs, si des erreurs sont présentes elles pourraient être corrigées relativement facilement. Nos données tendent à indiquer qu'un correctif s'avère nécessaire, la preuve n'en est cependant pas concluante. Le taux de productivité fourni par les firmes de sondages ou les autres firmes spécialisées en la matière sont des moyennes. Peut-être avons-nous obtenu une valeur extrême du taux et que, pour une autre enquête, celui-ci sera de 2 % plus élevé que celui publicisé.

Dans tous les cas, nous pensons avoir montré qu'une validation, la plus complète possible, de la banque de gestion d'appel est un passage obligé en vue d'obtenir des données fiables.

## RÉFÉRENCES

- Fourmier, L.; Chevalier, S.; Lemoine, O.; Poulin, C.; Poirier, L.-R. (sous presse) *La santé mentale des Montréalais. Rapport d'enquête*. Montréal : Régie régionale de la santé et des services sociaux de Montréal-Centre, Direction de la santé publique.
- Groves, R.M. (1989). *Survey errors and survey costs*. New York : John Wiley & Sons.
- Lavrakas, P.J. (1993). *Telephone survey methods*. Newbury Park : Sage.