

Les sondages Internet, défis nouveaux et revisités

Claire Durand¹

Résumé

L'article présente l'état des lieux concernant les sondages utilisant Internet, ceci sur deux plans, soit les questions relatives à l'échantillonnage et celles relatives au questionnaire. Il fait d'abord état des divers problèmes relatifs à la couverture, à la constitution de la base de sondage, à l'échantillon de départ, à la gestion des opérations et à la pondération. Il aborde ensuite les problèmes associés à l'élaboration du questionnaire. Suite à ces considérations, il trace un bilan et présente les perspectives d'avenir pour chacun des aspects mentionnés. Il se conclut sur le fait que, malgré les divers problèmes relevés, les sondages Internet sont des outils extrêmement flexibles et intéressants de collecte des données. Ils sont appelés à devenir le principal mode de collecte mais pour cela, certains problèmes devront être résolus.

Mots clés : Sondages internet ; méthodes de sondage ; méthodologie.

1. Pourquoi parler de sondages Internet

Internet est utilisé de plus en plus fréquemment comme mode d'administration des sondages et ceci non seulement pour les enquêtes marketing mais également pour les sondages sociopolitiques y compris les sondages d'intention de vote. Les recherches académiques ont également recours aux sondages faits par Internet et on retrouve des sondages « internationaux » utilisant Internet pour présenter un portrait de l'opinion publique de pays en développement où l'accès à Internet n'est pas très répandu. Internet est également utilisé comme mode d'administration principal ou complémentaire pour certaines enquêtes des agences statistiques. Ainsi, le Recensement canadien et l'Enquête nationale auprès des ménages menés par Statistique Canada en 2011 utilisent Internet comme mode d'administration complémentaire. Il est donc très pertinent d'examiner les défis que présente ce nouveau mode d'administration.

Pourquoi utilise-t-on Internet pour faire des sondages? D'une part, l'accès à Internet s'accroît rapidement, y compris dans les pays en développement et ce, d'autant plus que l'accès sans fil et mobile ne nécessite pas de construire des infrastructures coûteuses. Ainsi, la pénétration d'Internet en 2012 selon Internet World Stats (<http://www.internetworldstats.com/>, novembre 2013) est au-dessus de 75 pour cent dans les principaux pays européens de même qu'au Japon et en Corée du Sud. Elle est de moins de 15 pour cent en Algérie, de 34 pour cent au Vietnam, de 36 pour cent au Mexique mais de 51 pour cent au Maroc et de 66 pour cent en Argentine. D'autre part, l'administration de questionnaire via Internet est très peu coûteuse; il s'agit probablement du mode le moins coûteux disponible actuellement. Il permet d'économiser les coûts du papier, de la poste, des interviewers, des télécommunications ou du transport associés aux autres modes d'administration. Enfin, les sondages utilisant Internet comportent certains des avantages des sondages auto-administrés, notamment une moins forte pression à la conformité sociale, ainsi que des avantages des sondages face à face, notamment la possibilité de recourir à des présentations de matériel vidéo.

Le développement rapide des technologies de l'information et des communications (TIC) demande de faire le point régulièrement sur le développement des sondages Internet et sur les principaux défis qu'ils posent. Le texte présente les principaux défis méthodologiques classiques et les nouveaux défis posés par les sondages utilisant Internet puis conclut en faisant un état des lieux et en ouvrant sur les perspectives d'avenir pour chacun de ces défis.

¹Claire Durand, département de sociologie, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada, H3C 3J7, Claire.Durand@umontreal.ca.

2. Qu'est-ce qu'un sondage Internet?

Très simplement, un sondage Internet est un sondage effectué via un site Web. Toutefois, cette appellation générique cache de multiples manières de procéder pour recruter les répondants. Cooper (2000) présente huit modes de recrutement, dont trois sont non probabilistes et cinq probabilistes. Les modes non probabilistes sont les *vox pop* sur les sites des médias, le recrutement de volontaires pour une enquête unique et le recrutement de volontaires pour adhérer à un panel qui servira ensuite à de nombreuses enquêtes. Pour ces deux derniers modes, les participants sont habituellement « récompensés » de diverses manières incluant l'accès à des cadeaux de nature monétaire ou autre et la possibilité de gagner divers prix et. Ces récompenses sont rarement très élevées, ce qui respecte les codes d'éthique des associations regroupant les professionnels en recherche (ARIM, AAPOR, etc.). Toutefois, ils sont souvent utilisés comme argument pour le recrutement dans les panels. Pour ce qui est des cinq modes de sélection probabilistes, ils comprennent les sondages de sortie de site (où les visiteurs d'un site web sont interceptés au hasard pour répondre à un questionnaire), le recours aux listes courriel d'organisations ou d'associations (pour des sondages auprès des employés ou des membres), les sondages multi-modes avec possibilité d'utiliser Internet, les panels d'utilisateurs d'Internet recrutés par un autre mode (téléphonique, face à face, poste) représentant uniquement les utilisateurs d'Internet et enfin, les sondages visant à représenter l'ensemble de la population où les participants sont recrutés par un autre mode et où on fournit l'accès à Internet aux personnes sélectionnées qui ne l'ont pas. Seul ce dernier mode permet de constituer une base de sondage probabiliste de l'ensemble de la population. Toutefois, comme ce dernier mode est très coûteux, il en découle que la très grande majorité des sondages Internet visant à représenter l'ensemble de la population utilisent en ce moment des panels de volontaires recrutés au moins en partie de façon non probabiliste.

3. Défis méthodologiques classiques

Lorsque l'on aborde les erreurs des sondages, on parle habituellement d'erreur totale (Tourangeau, Conrad et Couper, 2013), répartie en deux catégories soit les erreurs de non observation liées à l'absence de certaines personnes dans l'échantillon des répondants et les erreurs d'observation liées à l'outil de collecte. Cette section porte sur les erreurs de non observation, c'est-à-dire celles qui sont liées à la couverture, à l'échantillonnage et à la non-réponse.

3.1 La couverture, l'exclusion

L'accès à Internet est évidemment « la » question lorsque l'on examine la couverture et l'exclusion. D'une part, cet accès est plus ou moins important selon les pays. Il se situe en ce moment autour de 80 pourcent dans la plupart des pays développés mais l'accès est nettement moindre dans les autres pays. Or, il existe des différences selon l'accès à Internet quant à la résidence, à la scolarité, à l'âge et au style de vie. Ainsi Bigot, Crouette et Recours (2010) comparent internautes et non internautes répondant au sondage « Conditions de vie et aspirations des Français », un sondage probabiliste en face à face réalisé par le Centre de Recherche pour l'Étude et l'Observation des Conditions de Vie (CREDOC). Ils trouvent des différences significatives avant redressement sur 71 des 191 variables (37%) dont celles portant sur l'ensemble des caractéristiques sociodémographiques à l'exception du genre. Après redressement, les différences persistent pour 12 pour cent des variables dont celles portant sur les pratiques culturelles et sur les opinions en matière de mœurs. On pourrait penser que si on constitue un échantillon probabiliste en fournissant l'accès à Internet à ceux qui ne l'ont pas, on élimine ce problème. Toutefois, encore faut-il que les personnes sélectionnées soient en mesure d'apprendre à se servir d'Internet. Un certain niveau de littératie est nécessaire pour répondre à tous les sondages auto-administrés mais pour répondre à un sondage Internet, il faut également un bon niveau de littératie technologique. La population avec accès à Internet et en mesure de s'en servir pour répondre à un sondage est donc un sous-ensemble de la population et ce sous-ensemble a des caractéristiques distinctives.

3.2 La constitution de la base de sondage

Une fois définie la population, il faut tenter de constituer une base de sondage qui la couvre le mieux possible. Pour les sondages téléphoniques et en face à face, cette base est constituée de numéros de téléphone et d'adresses correspondant à des ménages. Il est possible, plus ou moins facilement selon les pays, de constituer des bases de sondage correspondant assez bien à l'ensemble de ces ménages. Les échantillons doivent donc être à deux niveaux

puisque l'on devra d'abord sélectionner au hasard un ménage dans la base puis une personne au sein du ménage sélectionné. En principe, seules les personnes n'ayant pas le téléphone, n'ayant pas de domicile fixe ou ne vivant pas dans un ménage privé sont exclues de ces bases. Pour ce qui est des sondages Internet, on fait face à des différences majeures. D'une part, l'adresse électronique est généralement attachée à une seule personne plutôt qu'à un ménage. D'autre part, il n'existe pas de liste des adresses électroniques et il est pratiquement impossible d'en établir une. Pour constituer une base de sondage, certains sondeurs ont d'abord procédé de façon indirecte en demandant l'adresse courriel des répondants à leurs sondages téléphoniques et en les contactant ensuite pour les inviter à faire partie du panel. Il s'agit d'une méthode probabiliste apte à constituer éventuellement un échantillon d'internautes. On aurait pu penser que les sondeurs tenteraient ainsi de constituer petit à petit une base des adresses électroniques de la population. Toutefois, ce n'est pas ce qui s'est passé. La plupart des sondeurs et des pourvoyeurs de listes ont combiné plusieurs méthodes et ont eu recours principalement au recrutement de volontaires en plaçant des annonces sur leur propre site web ainsi que sur des sites Web externes, sur des moteurs de recherche et sur des média sociaux. Au final, il y a une grande variété dans la qualité des bases de panels de volontaires. Certains panels comprennent des personnes recrutées à la fois de façon probabiliste et non-probabiliste et certains sondeurs utilisent uniquement la partie probabiliste de leur base de panel pour certains types de sondage. D'autres ont recours à des panels constitués exclusivement de volontaires et les mêmes fournisseurs de listes peuvent fournir plusieurs sondeurs. D'autres, plus rares, utilisent seulement un recrutement probabiliste et généralisent leurs résultats uniquement aux internautes. Quoiqu'il en soit, il semble y avoir une « limite » à l'accumulation de membres dans les bases de panel puisque tout indique que les sondeurs ont cessé de faire grossir leur base de sondage et se contentent de renouveler ceux qu'ils perdent par attrition. « Entretenir » la base demande sans doute un certain suivi et entraîne donc des coûts.

Quelque soit le moyen utilisé pour les recruter, les personnes contactées sont invitées par la suite à compléter un questionnaire, parfois très détaillé, portant sur leurs coordonnées, leurs caractéristiques, leurs habitudes de vie et de consommation, etc. Cette étape, qui permet de s'assurer de l'identité des répondants et qui permet de cibler certains marchés pour les sondages marketing, peut constituer un frein pour certaines personnes, particulièrement dans les démocraties en émergence. Une fois toutes ces informations révélées, les répondants ont-ils toujours confiance que la confidentialité de leurs réponses sera respectée (Lozar Manfreda et coll. 1998)? Certains répondants décident-ils de ne pas s'inscrire parce qu'ils sont rebutés par la quantité d'informations précises qu'ils sont invités à fournir? Bref, cette étape peut entraîner des biais d'exclusion supplémentaires.

3.3 La constitution de l'échantillon de départ

Une fois le panel constitué, il faut procéder à la constitution de l'échantillon de départ pour chaque sondage. Les bonnes pratiques voudraient que l'on sélectionne l'échantillon « représentatif » initial le plus petit possible et que l'on tente de convaincre le plus de membres possible de cet échantillon de compléter le sondage. Toutefois, la plupart des sondeurs qui ont recours à des panels Internet de volontaires utilisent plutôt la méthode des quotas. Ils lancent un nombre très important d'invitations (jusqu'à plus de 100 fois le nombre de répondants souhaité) et arrêtent d'accepter les répondants au fur et à mesure que les quotas -- définis en général par l'âge et le sexe et parfois par la région et la langue -- sont remplis.

3.4 La gestion des opérations, le taux de réponse

Lozar Manfreda et coll. (2008), effectuant une méta-analyse de 45 expériences comparant le mode Internet à d'autres modes, montrent que les taux de réponse aux sondages Internet sont plus bas de 11 points en moyenne que ceux des autres modes. Pour ce qui est des panels de volontaires, on parle de taux aussi bas que moins de un pour cent. Le faible coût lié à l'augmentation du nombre de requêtes semble avoir amené certains sondeurs et chercheurs à privilégier le nombre de répondants aux dépens de leur représentativité. Ces bas taux de complétion résultent d'une manière de gérer les échantillons qui risque d'augmenter les biais de couverture. En agissant ainsi, on privilégie les personnes qui répondent plus rapidement et celles-ci n'ont pas nécessairement le même profil socioculturel et sociopolitique que les personnes « occupées ». De plus, cette manière de procéder a un effet boule de neige. Elle entraîne une augmentation du nombre de requêtes reçues par les membres et ceci d'autant plus lorsque ces derniers appartiennent à des groupes sous-représentés dans les bases de sondage. Ces membres finissent par être excédés par le grand nombre de requêtes qu'ils reçoivent et ne prennent plus la peine de répondre ce qui provoque une nouvelle hausse du nombre de requêtes nécessaires pour atteindre la taille requise de l'échantillon final. Il s'ensuit automatiquement une nouvelle baisse du taux de collaboration. Au final, on peut se retrouver avec des échantillons de répondants quasi professionnels, attirés par les possibilités de rémunération ou par le sujet de chaque enquête, plus

souvent chômeurs, célibataires, sans enfant. Or, cette manière de faire n'est pas inéluctable. On pourrait atteindre des taux de collaboration acceptables en utilisant les méthodes connues, développées pour les sondages auto-administrés: gestion serrée, rappels variés, *etc.* (Dillman, 2000). Certains sondeurs y parviennent en y consacrant les ressources nécessaires.

Il demeure qu'il faudra mieux comprendre Internet comme mode de communication. Tout concorde à montrer, par exemple, que les rappels par courriel fonctionnent moins bien que ceux utilisant la poste ou le téléphone (Tourangeau et coll., 2013). Des expériences montrent que des rappels utilisant des cartes postales, par exemple, seraient particulièrement efficaces. Le courriel est un média presque « instantané » : Alors que la carte postale peut être laissée à la vue pour servir de rappel, les courriels disparaissent de la vue très rapidement, surtout chez les grands utilisateurs. Or, les logiciels de gestion des courriels courants ne permettent pas de mettre facilement en exergue les courriels pour qu'ils se rappellent au souvenir du répondant.

3.5 Les correctifs possibles: la pondération et le redressement

Les chercheurs ont tenté de trouver des moyens de s'assurer que les sondages Internet utilisant des panels de volontaires représentent l'ensemble de la population ou à tout le moins se rapprochent d'une bonne représentation en utilisant diverses méthodes de redressement et de pairage. Les redressements utilisent les scores de propension basés sur la propension d'accès à Internet et diverses variables ayant trait entre autres au style de vie. Tourangeau et coll. (2013) dans une synthèse de huit études qui ont comparé diverses méthodes de redressement visant à réduire les biais des sondages Internet dont quatre portant sur des panels de volontaires, concluent que les redressements corrigent au plus 60 pour cent des biais, qu'ils augmentent parfois les biais jusqu'à un facteur de plus de deux, que les estimations peuvent être modifiées d'autant que 20 points et qu'il y a d'importantes différences selon les variables, les ajustements ayant pour effet tantôt d'éliminer les biais, tantôt de les augmenter sensiblement. De même, Loosveldt et Sonck (2008) ont montré que la pondération par score de propension permet d'ajuster pour la proportion d'urbains et de personnes en emploi mais que les différences demeurent pour ce qui est de la satisfaction face à l'emploi, de l'intérêt pour la politique et des attitudes face aux immigrants en Flandre belge.

3.6 En résumé, pour ce qui est des défis classiques

Les manières de faire problématiques se combinent et contribuent conjointement à produire des échantillons biaisés. Plusieurs recherches le montrent. Blasius et Brandt (2010) tentent de constituer un échantillon de panel représentatif du recensement allemand selon l'âge, le sexe et le niveau de scolarité. Ils réussissent à constituer un tel échantillon uniquement pour les 18 à 49 ans puisque, au-delà de cet âge, il n'y a pas assez de membres. La comparaison de cet échantillon avec deux autres échantillons provenant du recensement et du General Social Survey (GSS) montre que l'échantillon issu du panel Internet comprend proportionnellement plus de personnes célibataires ou divorcées, moins de personnes fréquentant l'église et, sur le plan des valeurs, plus de valorisation du laisser faire, de la richesse et de la réalisation de soi. De même, Stephenson et Crête (2011) comparent deux sondages faits au Québec en 2007 par la même firme avec le même questionnaire, un par téléphone, l'autre auprès d'un échantillon de panel Internet de volontaires. Ils montrent que 36 des 52 variables (69%) ont une distribution significativement différente, y compris après pondération. Comme Blasius et Brandt (2010), ils constatent la moindre présence de personnes pratiquant une religion dans l'échantillon Web et pour ce qui est des opinions, une plus grande propension à penser que « l'on est allé trop loin pour accommoder les minorités culturelles au Québec ».

Krosnick et ses collègues (Malhotra et Krosnick, 2007; Pasek et Krosnick, 2010) font des constatations similaires: Les sondages Internet comprennent moins de Noirs, moins de personnes peu scolarisées, plus de partisans de la guerre en Irak et plus de partisans de Bush que de Kerry en 2004. De plus, les relations entre les variables prédictives et le vote diffèrent (Malhotra et Krosnick, 2007). Comparées à un sondage téléphonique de type RDD, les proportions de réponses modales diffèrent en moyenne de 13 points et les différences vont jusqu'à 30 points. De même, des différences significatives apparaissent dans la prédiction de la participation au recensement, dans l'évolution des opinions avec le temps et dans les relations entre les variables (Pasek et Krosnick, 2010). Enfin, Durand (2012, 2013) montre que les sondages Internet diffèrent presque toujours des autres sondages dans la prédiction des résultats électoraux au Canada.

Bref, tout amène à conclure qu'il existe des différences non insignifiantes et parfois importantes entre les sondages Internet, particulièrement ceux qui utilisent des panels de volontaires, et les autres sondages utilisant des échantillons

probabilistes. Ces différences apparaissent non seulement pour ce qui est des caractéristiques démographiques, mais également, même après avoir contrôlé pour les différences relatives à la démographie, sur les valeurs, les opinions et les attitudes. Ces différences sont généralement non constantes et non systématiques. Pour ce qui est des panels Internet de volontaires, le biais relatif au recours à des volontaires s'ajoute au biais provenant de l'exclusion de ceux qui n'ont pas accès à Internet et il existe des différences relativement constantes quant à la pratique religieuse – moins importante -- et quant aux attitudes face à la diversité – plus négatives. Lorsque des quotas sont appliqués avec succès, on constate des problèmes similaires à ceux d'autres échantillons par quotas soit le fait que le respect des quotas n'assure pas une représentation sociopolitique adéquate (Durand, Blais et Larochelle, 2005; Curtice et Sparrow, 1997). Enfin, les redressements réduisent parfois mais éliminent rarement les biais dus à la couverture, à la sélection et à la non-réponse inhérents aux panels de volontaires.

4. Les questions et le questionnaire: de nouveaux défis

Les recherches sur la formulation des questions et la présentation de celles-ci sont nombreuses et beaucoup d'avancées ont été faites dans le cas des sondages Internet (Tourangeau et coll. 2013). Certaines recommandations des recherches ont été intégrées dans la plupart des logiciels. A titre d'exemple, les questionnaires web donnent généralement des indications sur la progression du questionnaire et permettent de revenir en arrière; des formats tableau sont utilisés pour les listes d'éléments ayant le même choix de réponse. De plus, la plupart des firmes rejettent automatiquement les questionnaires répondus trop rapidement. Toutefois, il arrive encore fréquemment que les questionnaires ne permettent pas la non-réponse aux questions.

Un questionnaire n'est pas une simple liste de questions. Il s'agit d'un outil global, à la fois un instrument de mesure et un mode d'interaction sociale. Contrairement à ce qui se passe avec les autres sondages auto-administrés, les questionnaires de sondage Internet sont présentés de façon fragmentée, soit question par question, page par page ou en défilement. Ils sont tributaires du fait que les premiers logiciels se sont développés pour une utilisation par les interviewers et non par les répondants. Le répondant n'a pas de vision d'ensemble du questionnaire. Il perd de vue ce qui précède au fur et à mesure qu'il répond aux questions et n'a pas idée de ce qui suit. Or la lecture sur écran comporte des difficultés inhérentes, différentes de la lecture sur papier (Jabr, 2013). Alors que les applications de lecture numérique se sont développées pour répliquer assez fidèlement la lecture sur papier, les logiciels de questionnaire Web n'ont pour la plupart pas encore atteint ce stade. On peut penser par exemple à l'intérêt qu'il y aurait à avoir des bandeaux verticaux à la gauche de l'écran qui permettraient au répondant de suivre où il en est dans le questionnaire et de naviguer plus facilement (Tourangeau et coll., 2013). De plus, certains questionnaires sont beaucoup trop longs et la lecture sur écran est plus susceptible d'entraîner de la fatigue visuelle (Jabr, 2013). Là encore, le faible coût des sondages fait que les chercheurs ont moins d'incitatifs à se restreindre. Ceci entraîne des taux d'interruption importants, d'autant plus importants que les questionnaires sont longs (Tourangeau et coll., 2013).

5. État des lieux et perspectives d'avenir

Quel bilan pouvons-nous dresser des divers défis mentionnés et quelles sont les perspectives d'avenir pour relever ces défis?

Pour ce qui est de la couverture, il est évident qu'Internet est en train de devenir « LE » mode de communication. Bientôt, l'accès sera à ce point généralisé, du moins dans les pays les plus développés que les problèmes d'exclusion pourront être considérés comme négligeables. On peut penser que les avancées technologiques vont continuer à faciliter l'accès à Internet et vont éventuellement le rendre disponible à tous à moindre coût. Dans cet éventualité, l'adresse courriel est appelée à devenir une adresse unique à chaque personne comme l'adresse postale ou le numéro de téléphone fixe jusqu'à tout récemment. À partir de ce moment, les problèmes de couverture seront en pratique résolus...mais pas les problèmes associés à la constitution d'une base.

Les seuls moyens disponibles pour constituer une base de sondage des adresses courriel de l'ensemble de la population sont trop coûteux pour être accessibles à la plupart des utilisateurs et commanditaires potentiels. Ce n'est donc pas une solution en ce moment ce qui amène les utilisateurs à se tourner vers les panels de volontaires. Par

contre, ceci a encouragé le développement de méthodes de recrutement non probabilistes diversifiées. De plus, on assiste au développement de méthodes probabilistes basées sur le recrutement téléphonique ou postal. La tâche à réaliser est la constitution de listes d'adresses courriel de l'ensemble des populations permettant de tirer des échantillons. Il s'agit d'une tâche difficile qui se heurtera à la nécessité de respecter le droit à la protection de la vie privée, un droit de plus en plus balisé. La nouvelle loi anti-pourriel canadienne, par exemple, empêchera d'envoyer des courriels sans avoir obtenu la permission préalable des récipiendaires et elle interdira de recueillir les adresses courriel sur Internet. Entretemps, les bases existantes devraient être développées pour inclure le plus possible de nouveaux « membres » via un recrutement téléphonique ou un autre mode probabiliste. Enfin, le plein potentiel des bases de type panel devrait être utilisé pour mieux comprendre les profils des volontaires et les profils de non réponse. Il demeure que, dans un tel environnement, il faudra sans doute prendre des moyens pour mieux informer la population des avantages de la recherche et aller chercher son adhésion et sa collaboration.

Pour constituer les échantillons de départ à partir des bases pour panel, on semble être retourné au mode de fonctionnement propre à l'utilisation de la méthode des quotas soit des échantillons beaucoup plus importants que nécessaire et donc un nombre d'invitations très élevé auprès des membres et encore plus auprès de ceux ayant des caractéristiques sous-représentées dans la base de panel. De plus, le recours à des incitatifs monétaires est devenu la norme avec les biais possibles que cela entraîne. Même si les incitatifs ne sont pas importants, il demeure que les publicités visant à recruter des membres font de cet aspect leur argument de « vente » principal. Revenir à des méthodes fiables connues de gestion des opérations – des échantillons fermés et des méthodes de maximisation du taux de réponse tels des rappels diversifiés -- serait un progrès important et très facile à réaliser. Rien n'oblige à fonctionner comme cela se fait en ce moment alors que l'impact de cette manière de procéder sur les biais des échantillons de répondants est possiblement important.

En ce qui a trait au redressement des échantillons, beaucoup de recherches ont été faites sur l'efficacité des méthodes basées sur le tirage, la probabilité d'inclusion (scores de propension) ou le recours à des variables externes. Ces tentatives ont donné des résultats plutôt décevants, variables et non prévisibles. Existe-t-il un ou des moyens de s'assurer qu'un échantillon de volontaires est représentatif de la population dans toutes les circonstances? Ou dans certaines circonstances? Ou d'en arriver à des biais prévisibles? La réponse simple à cette question est non. En conséquence, il semblerait plus productif de concentrer la recherche sur la constitution d'échantillons probabilistes et sur la pondération des échantillons probabilistes d'utilisateurs d'Internet. Cela permettrait de mieux comprendre s'il est possible -- et dans quelles circonstances -- de généraliser à la population à partir de tels échantillons en utilisant les informations récoltées sur les différences entre les utilisateurs et les non-utilisateurs d'Internet dans les échantillons probabilistes.

Enfin, pour ce qui est des questionnaires, beaucoup de recherches ont déjà été faites sur le format de la page-écran, sur les indicateurs de progression, sur les indications aux répondants de même que sur l'impact de permettre ou non la non réponse aux questions. Toutefois, pour ce qui est de la présentation globale du questionnaire comme instrument, peu d'avancées ont été faites au cours des dernières années, entre autres pour permettre au répondant d'avoir une vue globale du questionnaire ou même d'y répondre comme s'il était en format papier. Ce type de format est disponible pour les livres numériques et pourrait sans doute être implanté pour les questionnaires. Le répondant pourrait, au choix, décider de compléter le questionnaire comme s'il feuilletait un format papier, en sachant le nombre de pages, etc. Une autre option est celle du bandeau à gauche permettant d'avoir une idée plus globale de la progression du questionnaire. De plus, il y aurait lieu de profiter du fait que les questions ouvertes sont beaucoup plus faciles à traiter en mode électronique. Les principaux obstacles à leur utilisation dans les autres modes -- la nécessité pour l'interviewer de prendre note de la réponse et la nécessité de déchiffrer et transcrire les réponses -- sont absents dans les sondages Internet. Le répondant pourrait apprécier cette possibilité de présenter son opinion de façon moins balisée et la compréhension des réponses gagnerait en profondeur. Enfin, pour ce qui est de la recherche, il reste des balises à développer sur la longueur maximale des listes de questions et des questionnaires en ligne de façon à faire diminuer les taux d'interruption et à donner des indications aux chercheurs et aux sondeurs sur les limites à ne pas dépasser.

6. Conclusion

Le portrait tracé de la situation des sondages Internet peut apparaître négatif. Toutefois, il faut bien comprendre que ce mode d'administration est assez récent, que les développements dans les modes de communication se sont faits

très rapidement et que les divers acteurs ont donc dû s'adapter en accéléré. Il demeure qu'Internet est en voie de devenir le mode principal de collecte de données dans les pays industrialisés et dans les démocraties émergentes. Il s'agit d'un mode flexible qui comporte des avantages importants, dont le fait de pouvoir présenter du matériel vidéo, de recourir à des questions ouvertes, de permettre aux répondants de compléter le questionnaire à leur convenance sans les déranger ou leur imposer un horaire spécifique. Toutefois, il deviendra l'outil de choix si et seulement si certains problèmes de base sont résolus. L'absence de base de sondage fiable est sans doute le problème le plus criant et les coûts pour faire un sondage avec un échantillon probabiliste sont en ce moment prohibitifs. La manière de procéder pour gérer les opérations entraîne des taux de réponse inacceptables et alourdit le fardeau des membres de panel. Les tentatives pour ajuster les échantillons après coup n'ont pas donné les résultats espérés. Le questionnaire n'est pas encore un instrument global et convivial. Enfin, il demeure que, comme pour les sondages en face à face, puis téléphoniques, à leurs débuts, il faudra que la population « apprenne » à répondre à des sondages Internet. Il reste beaucoup de travail à faire pour convier l'importance de collaborer.

Bibliographie

- Baker et coll. (2013), Report of The AAPOR Task Force on Non-probability Sampling, <http://www.aapor.org/AM/Template.cfm?Section=Reports1&Template=/CM/ContentDisplay.cfm&ContentID=6055>, 128 p.
- Bigot, R., Croutte, P. et F. Recours (2010), Enquêtes en ligne : peut-on extrapoler les comportements et les opinions des internautes à la population générale? Centre de recherche pour l'étude et l'observation des conditions de vie (CREDOC), Cahier de recherche no. C273, 63 p., <http://www.credoc.fr/publications/abstract.php?ref=C273>.
- Blasius et Brandt (2010), Representativeness in Online Surveys Through Stratified Samples, *BMS*, 107, p. 5-21.
- Curtice, J. and N. Sparrow (1997), How Accurate Are Traditional Quota Opinion Polls? *Journal of the Market Research Society*, 39, 433-448.
- Dillman, D. (2000), *Mail and Internet Surveys, the Tailored Design Method*, New York: Wiley and Sons, 463 p.
- Durand, C., Blais, A. et M. Larochelle (2005), The Polls of The 2002 French Presidential Election: An Autopsy, *Public Opinion Quarterly*, 68 (4), 602-622.
- Durand, C. (2013), Why do Polls go Wrong Sometimes? The Canadian Case. Présenté à UBC Comparative Canadian and U.S. Politics Workshop, Department of Political Science, UBC, 23 septembre 2013, http://www.mapageweb.umontreal.ca/durandc/Recherche/Publications/pollsgowrong/Whydopolls_Canada.pdf.
- Durand, C. (2013), How do Internet Polls Fare in Predicting Election Results? Présenté au World Social Science Forum, Montréal, 15 octobre 2013, <http://www.mapageweb.umontreal.ca/durandc/Recherche/Publications/WSSF2013/howdointernetpollsfare.pdf>.
- Jabr, F (2013), Why the Brain Prefers Paper. *Scientific American*, November, p. 49-53.
- Loosveldt et Sonck (2008), An evaluation of the weighting procedures for an online access panel survey, *Survey Research Methods*, 2 (2), 93-105.
- Lozar Manfreda, K., Bosnjak, M., Berzelak, J. Haas, I. et V. Vehovar (2008), Web surveys versus other survey modes. A meta-analysis comparing response rates, *International Journal of Market Research*, 50 (1), 79-104.
- Malhotra, N et J. A. Krosnick (2007), The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples *Political Analysis*, 15(3), 286-323.

Pasek et Krosnik (2010), Measuring Intention to Participate and Participation in the 2010 Census and their Correlates and Trends: Comparison of RDD Telephone and Non-Probability Internet Survey Data, Statistical Research Division, US Census Bureau, Washington, Study Series Survey Methodology #2010-15, 71 p.

Stephenson et Crête (2011), Studying Political Behavior: A Comparison of Internet and Telephone Surveys, International Journal of Public Opinion Research, 23 (1), 24-49.

Tourangeau, R., Conrad, F. G. and M. P. Couper (2013), The Science of Web Surveys, New-York: Oxford University Press, 198 p.