



HAL
open science

La représentativité des échantillons issus d'access panels en ligne : une question majeure pour l'avenir des études de marché.

Ronan Divard

► To cite this version:

Ronan Divard. La représentativité des échantillons issus d'access panels en ligne : une question majeure pour l'avenir des études de marché.. 2009. hal-00819324

HAL Id: hal-00819324

<https://hal.univ-brest.fr/hal-00819324>

Preprint submitted on 30 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**La représentativité des échantillons issus d'access panels en ligne :
une question majeure pour l'avenir des études de marché**

Ronan DIVARD

Maître de conférences à l'Université de Bretagne Occidentale / Skol Veur Breizh Izel

Membre du laboratoire ICI (EA 2652)

IAE de Bretagne Occidentale, 12, rue de Kergoat – CS 93837 – 29238 Brest cedex 3
02 98 01 70 85, ronan.divard@univ-brest.fr

La représentativité des échantillons issus d'access panels en ligne : une question majeure pour l'avenir des études de marché

L'essor des access panels en ligne a bouleversé en profondeur le paysage des études de marché. Souvent présentés par leurs propriétaires comme des outils très performants, ils suscitent toutefois d'importantes interrogations scientifiques, particulièrement en ce qui concerne la représentativité des échantillons qui en sont issus et, partant, la validité des informations obtenues. Dans cet article, sont analysées toutes les sources de biais susceptibles d'affecter la représentativité des échantillons extraits d'access panels en ligne, ainsi que les moyens de les prévenir ou d'y remédier. Le but est notamment de permettre aux utilisateurs de tels dispositifs de disposer d'éléments les aidant à en apprécier la qualité.

Mots clefs: access panel, panéliste, enquête en ligne, erreur, échantillon, représentativité

The quality of online access panels: a major challenge for the future of market surveys

The development of online access panels deeply upsets the market surveys' landscape. Although they are presented by their owners as very successful tools, they often arouse important scientific questioning, particularly as regards the representativeness of the samples arising from them and therefore, the validity of the gathered information. In this paper, all the sources of error which can affect the representativeness of samples extracted from online access panels are analyzed, as well as the means to anticipate or to fix them. The main purpose is to provide the users of such devices with elements that will help them to evaluate their quality.

Keywords: access panel, panellist, online survey, error, sample, representativeness

La représentativité des échantillons issus d'access panels en ligne : une question majeure pour l'avenir des études de marché

Si les enquêtes conduites à partir d'access panels en ligne représentent une part croissante du marché des études, la qualité des données qui en sont issues et la validité même de la démarche suscitent de vives interrogations. Cet article vise à apporter des éléments de réponse à ces interrogations en analysant l'ensemble des erreurs susceptibles d'altérer la représentativité des échantillons issus d'access panels ainsi que les moyens de les prévenir ou d'y remédier. Un cadre d'analyse global spécifique aux access panels en ligne est proposé et des priorités de recherche sont énoncées.

L'article est articulé en quatre parties. Dans une première partie, nous traiterons des raisons qui expliquent la popularité des access panels en ligne ainsi que des interrogations qui entourent la qualité de ces dispositifs. Seront ensuite abordés logiquement les aspects relatifs à la constitution et la composition de l'access panel (erreur de couverture et erreur de recrutement), à son évolution (biais de conditionnement et biais d'attrition) et à la gestion d'une enquête particulière (erreur de sélection et erreur de non-réponse).

Une ascension qui s'accompagne d'interrogations

Un access panel en ligne peut être défini comme un vivier d'individus enregistrés qui ont volontairement et activement accepté de participer régulièrement à différents types d'enquêtes en ligne (17)¹. Il existe aussi des access panels hors ligne (panels postaux notamment), mais ils sont plus rares car ils sont bien plus coûteux et difficiles à gérer. Ils garantissent par contre *a priori* une meilleure représentativité, ne serait-ce que parce que l'inscription ne dépend pas d'un quelconque accès à Internet. Toutefois, il convient aujourd'hui de nuancer cette allégation, car il devient de plus en plus difficile, dans les panels postaux, de recruter des jeunes, particulièrement rétifs à ce mode de communication (21).

Les avantages des access panels en ligne

Les access panels en ligne permettent de renforcer les avantages habituellement reconnus aux enquêtes par Internet et de bénéficier d'atouts additionnels, ce qui explique leur fulgurante progression (2, 3, 16, 19) (voir encadré 1).

Encadré 1 – Les avantages spécifiques aux access panels en ligne

Un coût moindre : le pré-recrutement de répondants susceptibles d'être consultés lors de plusieurs enquêtes permet des économies substantielles ;

Des délais raccourcis : le pré-recrutement facilite la mobilisation rapide d'un échantillon et une collecte accélérée de l'information.

La possibilité de toucher des échantillons de grande taille : cette particularité permet notamment d'étudier avec précision de nombreux sous-ensembles de la population mère ;

La facilité pour ajuster rapidement la taille de l'échantillon : les relances sont peu onéreuses et, même si cette possibilité est rarement employée, il est loisible au chargé d'études de la reconsidérer en cours d'enquête, en fonction, par exemple, du taux de participation ou de la pénétration du produit étudié.

Une connaissance plus précise des répondants : lors du recrutement, sont collectées des informations sur des variables dites de qualification. Ces informations sont bien plus nombreuses que lors des enquêtes ponctuelles où on se contente habituellement de renseignements signalétiques sommaires. Elles permettent de tirer des échantillons ciblés ou stratifiés avec précision et facilitent les procédures de redressement ;

Un taux de réponse en principe élevé : les panélistes étant récompensés pour leur coopération, le taux de participation est normalement supérieur à celui des enquêtes *ad hoc*. L'existence d'un historique permet en outre d'anticiper sur le taux de réponses ;

Un contrôle des réponses facilité : les réponses aux différentes enquêtes peuvent être croisées et leur validité et leur fidélité plus aisément vérifiées ;

L'adaptation aux législations : les lois visant à protéger les internautes des intrusions rendent aujourd'hui bien plus malaisée la constitution d'échantillons *ad hoc* d'internautes. Disposer d'un vivier de volontaires permet de pallier cette difficulté ;

La détection des populations rares : la taille de la base et le volume d'informations détenues permettent de repérer et d'interroger des populations rares à un coût raisonnable ;

Un filtrage et un recrutement facilités pour les études qualitatives : il est possible de recruter plus aisément des personnes présentant le profil recherché pour des études qualitatives ;

La possibilité de conduire des études en continu : du fait de la taille et la relative pérennité du vivier, on peut mener des études longitudinales de type *tracking* et créer des panels dédiés.

L'essor des *access panels* en ligne s'explique également pour partie par la baisse générale des taux de participation aux enquêtes menées par des voies traditionnelles, qui a pour corollaires une augmentation des coûts et une dégradation de la valeur des estimations (1, 2, 3, 13). Ils seraient notamment une solution à l'érosion de l'efficacité des enquêtes téléphoniques qui se heurtent à des difficultés croissantes : réticences plus vives à délivrer de l'information et à donner de son temps, présence au domicile plus difficile à cerner, représentativité altérée par l'essor des numéros protégés et des « exclusifs mobile ».

La dégradation des taux de participation est souvent spectaculaire. Ainsi, aux Pays-Bas, le taux moyen de réponses aux enquêtes téléphoniques sur échantillons probabilistes a chuté de 60% à 30% pendant les années quatre-vingt-dix (3).

La qualité – un enjeu essentiel

L'association entre un faible coût relatif du terrain, la simplicité apparente des dispositifs et un taux de marge confortable a suscité un vif attrait pour le marché de l'*access panel* en ligne, d'autant plus que les barrières à l'entrée étaient basses, du fait de l'absence de standards clairs de qualité. De fait, le marché a été investi par des opérateurs qui n'avaient pas nécessairement une expertise en matière d'études ni une claire conscience des risques méthodologiques (1,7). En particulier, l'un des arguments commerciaux souvent mis en avant par les opérateurs – une relation implicite entre la taille de l'*access panel* et sa qualité – est des plus discutables. S'il est vrai qu'une trop petite taille ne permet pas de tirer pleinement partie des atouts potentiels d'un *access panel*, l'hypothèse implicite selon laquelle une grande taille se traduirait par une meilleure validité des estimations n'est pas théoriquement fondée. En effet, c'est simplement dans le cas des échantillons probabilistes que l'augmentation de la taille de l'échantillon améliore avec certitude la précision des estimations (2).

Ces dernières années, un large consensus s'est fait jour sur la nécessité de dépasser les simples considérations de performances quantitatives (effectif sondé, rapidité et coût du sondage) pour se livrer à un examen critique des méthodologies.

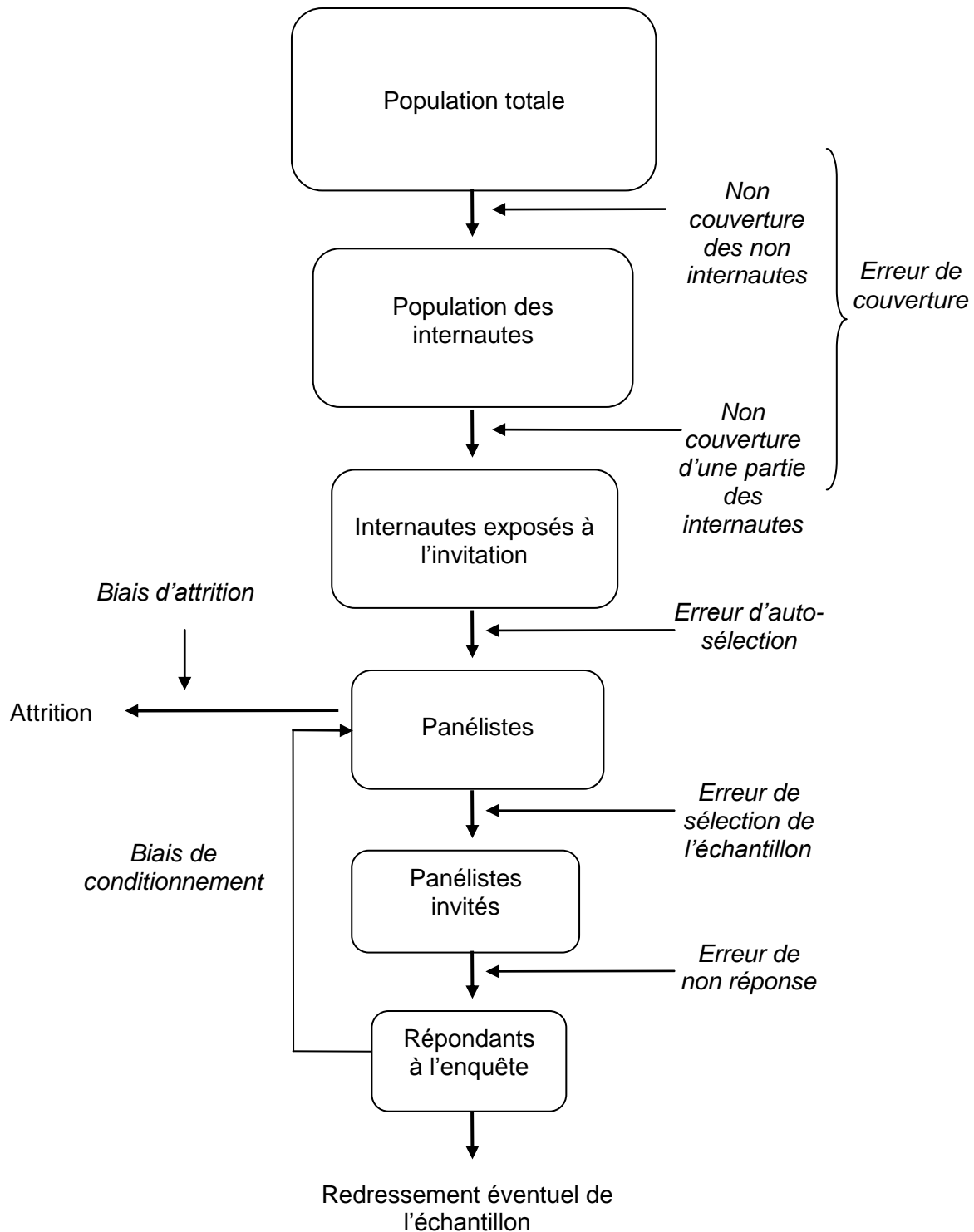
Non seulement les opérateurs les plus sérieux ont pris des initiatives destinées à remédier aux principales menaces (inscriptions multiples, identification des fraudeurs...) afin de rassurer

les sociétés clientes², mais les organismes professionnels ont édicté des règles visant à garantir la qualité scientifique des access panels et mieux réguler le marché.

En particulier, l'Esomar a conçu une série de 26 questions destinées à guider les clients tandis que d'autres organismes tels l'Efamro ou The Advertising Research Foundation proposaient des normes à portée universelle (1). Une norme ISO spécifique (ISO DIS 26362) a par ailleurs été publiée en 2008.

C'est dans le cadre de cette réflexion sur la qualité de ces dispositifs et les possibilités de l'améliorer que s'inscrit notre recherche. Elle est centrée sur la question fondamentale de la représentativité des échantillons issus d'access panels, qui correspond d'ailleurs à la principale réserve émise par les acheteurs d'études (4) et à une des exigences majeures énoncées dans la norme ISO, qui impose que les rapports d'étude mettent en évidence les limites à la transposition des résultats à la population mère définie. Cet article a donc pour ambition de mettre en évidence, dans un cadre d'analyse global, les différentes erreurs qui peuvent altérer la représentativité des échantillons obtenus à partir d'access panels en ligne ainsi que les moyens de les prévenir et d'y remédier (voir figure 1).

Figure 1 - Les sources d'erreurs susceptibles d'affecter la représentativité des enquêtes menées à partir d'access panels en ligne



La constitution et la composition de l'access panel

Il est possible à ce stade d'identifier deux sources d'erreur potentielle: l'erreur de couverture et l'erreur d'auto-sélection.

L'erreur de couverture est de deux types : l'absence de couverture des non internautes et l'absence de couverture d'une partie des internautes. L'absence de couverture des non internautes est inhérente au fait que, sauf exception, seules les personnes disposant déjà d'un accès Internet sont susceptibles de faire partie de l'*access* panel. Or, les internautes présentent des caractéristiques distinctives, qui peuvent affecter les résultats, dès lors que ceux-ci sont destinés à être extrapolés à une population plus large. L'absence de couverture d'une partie des internautes est imputable aux méthodes de recrutement utilisées qui, souvent, ne permettent de couvrir qu'une partie, généralement non représentative, de la population des internautes.

Quant à l'erreur d'auto-sélection, elle tient au fait qu'une fraction seulement des internautes donnera une suite favorable à l'invitation à devenir panélistes et que cette fraction peut présenter des spécificités susceptibles de porter atteinte à la validité des résultats.

L'absence de couverture des non internautes :

En toute logique, les résultats ne devraient pouvoir être inférés au mieux qu'à la population des internautes. Dans la pratique, ils sont souvent généralisés à une population mère plus large, ce qui pose un important problème de validité externe (2, 18). La parade est de procéder à une enquête sur une base probabiliste (une enquête téléphonique par *Random Digit Dialing* (RDD), par exemple) et de fournir un accès Internet aux répondants potentiels non équipés en échange de leur engagement à participer à une série d'enquêtes. Cette méthode est fort dispendieuse et le seul exemple connu est celui de *Knowledge Networks* qui recrute ses panélistes par téléphone et les équipe en accès Internet gratuit via la télévision. Les avantages d'une telle procédure sont indéniables : elle élargit la couverture aux non internautes, permet de mesurer les taux de non-réponse, autorise l'échantillonnage probabiliste et évite, du fait de la standardisation de l'équipement, les distorsions imputables aux problèmes techniques.

L'erreur de couverture des non internautes a pour origine un accès à Internet qui n'est pas généralisé et une distribution inégale de cet accès dans la population. Parmi les groupes sous-équipés, on compte en particulier les seniors et notamment les plus de 65 ans, les inactifs, les catégories sociales modestes, les ruraux, les minorités ethniques et les personnes à faible niveau d'études (1, 2, 21). Les distorsions peuvent être très importantes. Par exemple, en République tchèque, la pénétration d'Internet est de 46%, mais de 16% seulement chez les plus de 55 ans et de 6% chez ceux qui ont arrêté leurs études au plus tard à 15 ans (1).

L'assertion selon laquelle la croissance du nombre des internautes devrait réduire le biais de couverture est en partie discutable. En effet, l'ampleur du biais est déterminée par la proportion de non internautes, mais aussi par la différence entre internautes et non internautes sur les variables étudiées et on ne peut exclure que la minorité sans accès à Internet présente des caractéristiques comportementales et attitudeles de plus en plus distinctives, ce qui maintiendrait un biais substantiel (2).

Au-delà des simples considérations démographiques, il est important de comprendre les attitudes et comportements qui peuvent expliquer l'usage ou non d'Internet et de voir comment ils peuvent être liés aux sujets de recherche, par exemple à leur réceptivité face à certains produits ou services. S'il est plausible que des études portant sur des produits d'usage général tels que le dentifrice ou les céréales soient peu affectées, il n'en va sans doute pas de même pour des produits tels que les biens électroniques ou des comportements comme la lecture des quotidiens (1).

L'absence de couverture d'une partie des internautes

Si les méthodes de recrutement envisageables sont très nombreuses, la plupart d'entre elles ne donnent pas la même probabilité à chaque internaute d'être exposé à l'invitation à s'inscrire et ne permettent pas en outre de certifier l'identité des personnes (11) (encadré 2)

**Encadré 2 – Les principales méthodes de recrutement utilisées
pour les *access panels* en ligne**

Recrutement à partir d'un *access panel* existant (postal)

Recrutement à la suite d'enquêtes hors ligne (e. g. enquêtes omnibus)

Recrutement par téléphone, par contact personnel ou par courrier à partir de bases de données

Recrutement par diffusion de *flyers*

Recrutement par annonces presse

Recrutement par campagne d'*e-mailing* à partir de bases de données d'internautes

Recrutement à partir de bannières ou de *pop-ups* sur Internet

Recrutement par interception aléatoire

Recrutement à partir de liens avec des moteurs de recherche, des fournisseurs d'accès Internet, des sites marchands, des portails ...

Recrutement ouvert à partir du site de la société d'études

Schéma de recrutement par parrainage par les panélistes actuels

Sources : Postoaca A. (2006), *The anonymous elect – Market research through online access panels*, Berlin-Heidelberg, Springer; Comley P. (2007), Online market research, in *Market research handbook*, 5ème édition, Esomar World Research Publication, M. van Hamersveld et C. de Bont eds, John Wiley & Sons, 401-419.

Idéalement, il conviendrait de procéder à un recrutement aléatoire, ce qui est impossible, la liste exhaustive des internautes n'existant pas. Pour s'en approcher, on peut tirer un échantillon à partir d'une base de sondage probabiliste plus large. Cette méthode proactive permet d'éliminer le biais d'exposition et, en théorie, d'obtenir un échantillon représentatif des internautes, sous l'hypothèse d'une erreur de non réponse à l'invitation nulle ou insignifiante (2), hypothèse assez hasardeuse, comme nous le verrons.

Dans les faits, la plupart des *access panels* sont constitués à partir d'un recrutement qui n'a absolument rien de probabiliste (1, 2). En effet, les méthodes les plus fréquemment utilisées par les sociétés d'études, c'est-à-dire des invitations à s'inscrire en ligne, non seulement ne donnent pas à chaque internaute la possibilité de devenir panéliste, mais ne permettent aucunement de calculer la probabilité que chacun d'entre eux a de l'être. On peut parler à cet égard de biais d'exposition aux invitations (tous les internautes ne sont pas exposés à l'invitation à s'inscrire et le nombre d'expositions peut varier selon les internautes exposés).

A défaut de procéder à une sélection aléatoire, on peut limiter le biais en privilégiant un recrutement hors ligne, soit *ex nihilo*, soit en s'adressant aux participants à des enquêtes *ad hoc*. L'intérêt de cette approche est simple : le fait qu'un internaute soit sollicité ne dépend alors ni de l'intensité de sa fréquentation d'Internet, ni de sa propension à visiter tel ou tel site. Si l'institut décide de recruter en partie ou en totalité en ligne, il est éminemment recommandé de diversifier les sources afin de recruter une large variété de profils (5, 7, 21) et de ne procéder que sur invitation, soit en menant des campagnes d'*e-mailing*, soit en optant pour des messages à apparition aléatoire qui correspondent à un tirage systématique sur les visiteurs des sites concernés (16). Les méthodes permettant aux internautes de s'inscrire de

leur propre initiative doivent être prohibées (3). Car c'est bien au stade du recrutement que l'on peut prévenir l'intrusion des répondants professionnels en empêchant les démarches volontaires d'inscription (*via* le site de la société d'études, notamment). Il a été clairement montré à cet égard que l'auto-inscription amenait à des proportions de multi-inscrits et un nombre moyen d'appartenances bien plus élevés que lorsque l'opérateur recourait à des approches proactives basées sur des invitations (22).

Un biais d'auto-sélection qui peut être très marqué

L'inscription à un *access panel* étant basée sur le volontariat, les panélistes ne sont pas nécessairement représentatifs des internautes sollicités ni *a fortiori* des internautes en général. On peut seulement énoncer que la population est composée de personnes qui ont un accès à Internet, qui ont une probabilité non nulle d'être confrontées à l'invitation et qui décident de participer au panel (2). Les problèmes de représentativité des panélistes sont largement attestés (même s'il est souvent dans la pratique difficile de dissocier le biais d'exposition du biais d'acceptation). Les caractéristiques distinctives des internautes tendraient à être exacerbées: les panélistes sont notamment plus jeunes, actifs et éduqués que les internautes en général, les minorités ethniques sont très peu représentées et on trouve parmi les panélistes davantage d'utilisateurs réguliers et intenses d'Internet (5, 9, 18, 21, 22).

Le recours à une base probabiliste ne permet certes pas d'éliminer le biais d'auto-sélection. Par exemple, Burn et Thomas (4) obtiennent 19% d'acceptation à la suite d'une sollicitation par téléphone d'un échantillon probabiliste d'internautes et constatent que le taux d'acceptation est plus élevé parmi les moins de 35 ans, les étudiants et les inactifs et plus faible chez les retraités et les plus de 45 ans (il se dégrade au fil des classes d'âge). Même lorsqu'il existe des relations établies entre l'institut et les internautes, par exemple dans le cas de la migration d'un panel postal, la déperdition est très importante, les taux de transferts étant compris entre 25% et 50% seulement des panélistes postaux disposant d'un accès Internet. Ce qui impose aux instituts de recourir à des approches multimodales pour préserver la représentativité de leur panel. Van Walwyk et Garland estiment ainsi que, pour GfK Research Panels, il est nécessaire d'avoir 20% de panélistes hors ligne pour maintenir la validité des données (21).

En résumé, même si les panélistes sont recrutées activement par une autre voie qu'Internet et à partir d'une approche probabiliste, ceux qui sont intéressés et acceptent de s'inscrire

présenteront des particularités (5). Toutefois, il semble que le biais n'a pas un effet aussi négatif que dans le cas d'un échantillon non probabiliste. Ainsi, une étude comparative (11) montre que la qualité des données est bien meilleure quand on a recours à un échantillon probabiliste d'internautes. On observe en particulier de moindres proportions de répondants complétant hâtivement le questionnaire ou produisant des séries de réponses identiques dans des échelles d'attitude, ainsi qu'une amélioration de la cohérence des données d'enquête (e. g. la corrélation entre la satisfaction à l'égard d'une marque et la probabilité qu'elle se trouve dans l'ensemble de considération est sensiblement plus élevée).

Le premier déterminant du biais d'auto-sélection tient aux mobiles d'adhésion des panélistes (2). Il est fréquent en la matière de se référer à la théorie de l'autodétermination de Deci et Ryan (1975) en distinguant les motivations intrinsèques, qui conduisent à faire une action parce qu'elle est intéressante ou agréable, et les motivations extrinsèques qui amènent à l'accomplir pour des raisons externes (10). Le postulat est que les motivations intrinsèques doivent prédominer, car elles conduisent les panélistes à accomplir les tâches avec plus d'attention, de créativité et de plaisir (10, 16). Agir sur des motivations extrinsèques telles les récompenses monétaires peut même affecter les motivations intrinsèques. Il a ainsi été montré, lors d'une expérimentation conduite à l'aide du *Intrinsic Motivation Inventory (IMI)*, que le groupe récompensé était moins motivé intrinsèquement (10). S'il semble pertinent de renforcer les motivations intrinsèques en jouant sur différents facteurs (valeur de l'étude, valeur de l'opinion délivrée, autonomie du répondant, valorisation de sa compétence) (10), on notera toutefois que certains ressorts intrinsèques comme la curiosité sont susceptibles de biaiser dans le sens de l'optimisme les études sur des concepts ou des nouveaux produits.

Les motivations mises à jour sont très variées : possibilité d'améliorer la qualité des produits, curiosité, caractère divertissant des enquêtes, opportunité de s'exprimer sur de nombreux sujets, volonté de voir son opinion prise en compte, possibilité d'influencer les décisions, intérêt des sujets traités, perspective de gagner de l'argent ou des prix (17). Mais les études conduites sur la question ne permettent pas de dégager une hiérarchie nette des mobiles et de départager motivations intrinsèques et extrinsèques (6, 16, 22). Ces divergences peuvent tenir à la manière dont le recrutement a été mené et à la mentalité collective propre à chaque *access panel* étudié. On ne peut par ailleurs exclure l'existence d'un biais de désirabilité sociale amenant à minorer les mobiles intéressés. L'étude de Namiranian, Moskowitz et Gofman (14) apporte également un éclairage sur les motivations des panélistes, bien qu'elle ne porte pas

sur l'inscription initiale. Ces chercheurs ont mesuré à l'aide d'une analyse conjointe l'impact des arguments contenus dans une invitation à une enquête. Chaque panéliste devait classer une série d'invitations correspondant à différentes combinaisons d'arguments en indiquant dans quelle mesure il pensait y donner une suite favorable. Il en ressort que les récompenses jouent un rôle moteur (six des huit arguments qui ont la plus forte utilité y ont trait), loin devant l'intérêt du sujet et la durée du questionnaire.

Des segmentations des panélistes sur la base des avantages recherchés ont été proposées (encadré 3). Elles peuvent servir à différencier les politiques de communication et d'incitation ou à les mixer pour constituer un *package* propre à motiver le plus grand nombre (6).

Encadré 3 - les segmentations selon les avantages recherchés par les panélistes

Plusieurs segmentations à partir des bénéfices attendus par les panélistes ont été proposées. Elles aboutissent à distinguer quatre catégories de panélistes, qui diffèrent légèrement d'une analyse à l'autre.

Comley distingue les quatre segments suivants : les *Helpers*, qui aiment avant tout participer à des enquêtes et sont disposés à le faire sans rétribution, les *Opinionated*, particulièrement désireux que leurs opinions soient prises en compte, les *Incentivised*, très motivés par les loteries et les offres spéciales et les *Professionals*, surtout avides de récompenses monétaires directes.

Namiranian, Moskowitz et Gofman identifient également quatre segments : les *Survey Addicts*, largement prédisposés à participer, les *Give Me Stuff*, très attirés par les récompenses, les *Make It Easy And Convenient*, plus sensibles à la durée et à la flexibilité de la tâche et les *Helpful Expert Wannabee*, plus perméables à l'appel à l'altruisme et à l'invocation de leur expertise.

Van Walwyk et Garland, sur la base des études internes de GfK, distinguent quatre types de panélistes : les *Joiners*, qui sont heureux d'adhérer et de se sentir partie prenante de quelque chose d'important, les *Strong Opinion Holders*, qui sont convaincus d'avoir quelque chose à exprimer et qui veulent que leur opinion soit prise en compte, les *Incentive Seekers*, qui sont en quête de récompenses, qu'elle qu'en soit la nature et les *Helpful*, qui sont heureux de pouvoir apporter leur concours quand on les sollicite et qui ont souvent du mal à dire non.

Sources: Comley P. (2005), Understanding the online panellist, *Esomar Panel Research Conference*, Budapest; Namiranian L., Moskowitz H. R. et Gofman A., Motivations to join - What language do online panellists need to hear to volunteer? *Panel Research 2006, Esomar World Research Conference*, Barcelone; Van Walwyk M. et Garland C. (2008), Turning the super tanker – The migration from a postal to online methodology, *Panel Research 2008, Esomar World Research Conference*, Dublin

Les répondants professionnels : le Talon d'Achille des access panels en ligne ?

L'attrait pour les récompenses soulève le problème des répondants « professionnels » dont la présence est généralement jugée compromettante pour la valeur des études (16). L'appât du gain peut en effet les inciter à bâcler leurs réponses, à affirmer fallacieusement être éligible lors des filtrages, à s'inscrire sous plusieurs identités ou à devenir membre de plusieurs panels (1, 8, 16). En outre, leur participation à un nombre élevé d'études peut participer à un biais de conditionnement. Un autre problème est que l'appartenance multiple rend largement inopérantes les règles de management des *access panels*, les sociétés d'études n'ayant pas connaissance des autres activités de leurs panélistes multi-inscrits. L'inscription multiple serait un phénomène de grande ampleur. Selon une évaluation alarmante, un panéliste étasunien serait en moyenne inscrit à trois *access panels* et à sept s'il fait partie d'une population rare (16). *ComScore Networks* estimait en 2006 que 1% des internautes complèteraient 34% des questionnaires en ligne (4)³. Vonk, van Ossenbruggen et Willems (22), qui ont étudié 19 *access panels* néerlandais, évaluent le nombre moyen d'inscriptions à 2,73 avec de fortes variations (de 1,7 à 5,4). Encore s'agit-il de données déclaratives des panélistes... Les multi-inscrits tendent à présenter des caractéristiques distinctives. Il s'agit notamment d'une population plus féminine, plus jeune, et l'on y trouve davantage d'utilisateurs intenses d'Internet (22). Même si les études ne permettent pas de conclure formellement à une qualité de réponse intrinsèquement plus faible des multi-inscrits (22), il n'en demeure pas moins que la professionnalisation des répondants peut affecter gravement la confiance des clients et doit être combattue, de même que la professionnalisation de la participation aux *focus groups* l'a été par le passé. Et, comme nous l'avons souligné, c'est bien au stade du recrutement que l'on peut prévenir l'intrusion des répondants professionnels en rendant impossibles les démarches volontaires d'inscription.

Les facteurs qui influent sur le taux d'acceptation et le profil des panélistes

Si, d'une manière générale, il apparaît de plus en plus coûteux et difficile de recruter, plusieurs facteurs sur lesquels peut agir la société d'études sont susceptibles d'influer sur le taux d'inscription et le profil des personnes sollicitées.

1/ le mode de recrutement : son incidence sur le taux d'acceptation est nettement attestée. A titre d'exemple, Göritz (9) obtient à l'issue d'une expérimentation des taux d'accord de 1,3% pour les courriers, 7% pour les *flyers*, 7,7% pour les télécopies et 25,5% pour les courriels. Körner et Nimmergut (13) ont également comparé plusieurs méthodes (il s'agissait ici de recruter un *access panel* hors ligne pour un organisme public). Le taux d'acceptation est nettement plus élevé lors des présentations en face-à-face (18% pour les enquêteurs habituels et 26% pour des enquêteurs spécialement entraînés) que lorsque la documentation est simplement remise par l'enquêteur (8%) ou adressée par courrier (6%). La médiation d'un enquêteur semble donc décisive. Il s'agit surtout en fait d'exploiter les contacts lors des enquêtes *ad hoc* ou *omnibus*, car dépêcher spécifiquement des collaborateurs a un coût prohibitif. Mais cela suppose que la société d'études ait des activités de sondage classiques. Le mode de recrutement influe aussi sur la composition de l'*access panel*, non seulement sur le plan sociodémographique, mais aussi en termes d'usage d'Internet (9, 18).

2/ les récompenses initiales : elles peuvent faciliter le recrutement, l'impact étant supérieur si la récompense est garantie plutôt qu'aléatoire (loterie) (9).

3/ l'invitation et le questionnaire de qualification : l'invitation à s'inscrire a une influence, par les arguments qu'elle met en avant et les autres informations éventuellement apportées, sur le taux d'acceptation et la composition de l'*access panel* (16).

Nous sommes par ailleurs confrontés à un dilemme en ce qui concerne les informations qualifiantes à recueillir. D'un côté, plus elles sont nombreuses et précises, plus il sera aisé de cibler et stratifier des échantillons et plus on évitera de solliciter les panélistes à mauvais escient (6). D'un autre côté, le questionnaire de qualification peut être un frein à l'inscription si la tâche est perçue comme fastidieuse et indiscreète. On peut y remédier partiellement en récompensant la participation aux enquêtes de qualification et d'actualisation ou en échelonnant le recueil des données (16), même s'il est vrai qu'on peut douter de la participation aux enquêtes ultérieures de personnes qui seraient réticentes à compléter le questionnaire initial (6).

L'évolution de l'*access panel*

L'*access panel* peut voir sa représentativité s'éroder au fil du temps sous l'effet de deux phénomènes : le biais de conditionnement et le biais d'attrition.

Le biais de conditionnement

Le biais de conditionnement est un changement dans l'opinion ou le comportement mesuré chez un individu imputable au seul fait d'opérer la mesure. Il peut s'agir d'un changement réel ou d'un changement dans la manière de répondre. Il est en général jugé préjudiciable, bien qu'il puisse se traduire dans certains cas par une plus grande précision et une validité accrue des réponses. La plupart des études ayant porté sur des études longitudinales (des *trackings* de marques et de publicités) concluent à un biais de conditionnement faible ou insignifiant dans les *access panels* en ligne (5). Mais, il ne faut pas oublier que, même quand les enquêtes portent sur des sujets variés, le simple fait d'être panéliste peut générer un biais, biais qui peut être aggravé par l'appartenance multiple (22). Quelques différences intéressantes selon le degré d'expérience du panéliste ont pu être relevées : il apparaît notamment que les panélistes inexpérimentés consomment davantage d'informations, phénomène qui peut être attribué à la stimulation inhérente à leur inscription, qu'ils répondent plus positivement aux tests de concepts et qu'ils complètent plus lentement les questionnaires, ce qui pourrait s'expliquer par une moindre dextérité (5). La prudence recommande donc, comme le prescrivent les organismes professionnels, de ne pas solliciter les panélistes qui ont participé récemment à une enquête identique ou similaire et d'intégrer l'expérience des panélistes dans les critères de stratification lors du tirage des échantillons.

Le biais d'attrition

Le biais d'attrition dépend du taux d'érosion du vivier (l'attrition) et des différences entre ceux qui partent et ceux qui en restent membres. Si l'attrition est en général comprise entre 10 et 40% par an (7), il est difficile de porter un jugement sur l'existence d'un biais significatif, en l'absence d'études conduites sur cette question. De surcroît, la question de l'attrition et du biais qui peut en résulter est assez complexe. En effet, s'il semble *a priori* nécessaire de la limiter, non seulement parce qu'elle peut altérer la représentativité, mais aussi parce qu'elle impose de coûteux efforts de recrutement, l'attrition est parfois souhaitable, notamment parce qu'elle permet d'éliminer les panélistes indésirables. Il convient donc de bien analyser les sources d'attrition avant de mettre en place des règles d'éviction et une politique de rétention.

Les différentes formes d'attrition et leur traitement

L'attrition peut prendre quatre formes dans le cas particulier d'un *access panel* : la radiation des panélistes injoignables, les départs volontaires, l'éviction des panélistes inactifs et l'élimination des panélistes indésirables.

1/ Concernant les panélistes injoignables, il est conseillé de ne les évincer qu'après avoir tenté de résoudre le problème par téléphone.

2/ Pour les départs volontaires, la règle est d'éviter toute barrière à la sortie, le panéliste devant pouvoir se désengager de manière simple et rapide.

3/ Les organismes professionnels insistent sur la nécessité de ne comptabiliser que les panélistes actifs, c'est-à-dire ceux qui ont participé à au moins une enquête (même à un simple filtrage) au cours d'une période (une année selon l'Esomar) ou qui ont donné suite à au moins une sollicitation sur les x dernières adressées. La purge des panélistes inactifs n'est pas systématique. Certains *access panels* conservent ainsi des membres qui n'ont pas répondu depuis un an malgré de nombreuses sollicitations et, pire encore, certains opérateurs n'ont même pas d'informations détaillées sur l'historique de leurs membres... (22). Ces réticences à évincer les panélistes inactifs peuvent s'expliquer par la négligence, l'absence de moyens, mais aussi par le rôle essentiel que joue la taille de l'*access panel* dans l'argumentation de nombre de prestataires d'études (7).

4/ Les panélistes à problèmes ne constituent pas une population homogène. *Harris Interactive* les classifie ainsi en quatre catégories qui peuvent se recouper : les fraudeurs, les inattentifs, les hyperactifs et les conditionnés. Si les trois premières catégories doivent être considérées comme indésirables, les « conditionnés » nécessitent simplement une éviction temporaire et partielle. Plusieurs procédures de *scoring* ont été imaginées pour détecter les panélistes à problèmes à partir de l'analyse des questionnaires et de l'historique des répondants. Vonk, van Ossenbruggen et Willems (22) proposent ainsi d'évaluer les panélistes à partir de trois indices : l'indice de professionnalisme (nombre d'inscriptions à des *access panels* + nombre de questionnaires complétés au cours des douze derniers mois), l'indice de loyauté (temps consacré à répondre au questionnaire + longueur des réponses aux questions ouvertes) et l'indice d'inattention (indice de loyauté – 1). Les groupes caractérisés par des indices élevés (professionnels, loyaux, inattentifs) sont présents à des degrés très variables d'un *access panel* à l'autre et sont nettement atypiques sur des indicateurs comme la connaissance des marques et des publicités. Les proportions d'inattentifs et de professionnels sont en outre fortement

corrélées. L'explication pourrait tenir aux modes de recrutement. En effet, les panélistes loyaux sont proportionnellement plus nombreux quand l'institut recrute activement tandis que les pourcentages de panélistes inattentifs ou professionnels sont plus élevés quand l'opérateur recourt à des bases de données ou à l'auto-inscription.

Les axes d'une stratégie de rétention

Une stratégie de rétention efficace peut s'articuler autour de différents axes :

1/ Des récompenses pertinentes : un panachage d'avantages matériels et non matériels faciliterait la création d'une relation durable (7). Il serait également souhaitable de faire varier les récompenses dans le temps pour renouveler l'intérêt des panélistes (21).

2/ une véritable politique de communication et d'animation : Sassinot-Uny et Gadeib (17) jugent que le panéliste doit être considéré comme un client et l'appartenance au panel comme un service, ce qui impose d'améliorer les outils marketing destinés à conforter et pérenniser les relations avec les panélistes. De fait, la qualité de l'ensemble des supports de communication participe à la satisfaction des panélistes: invitation à s'inscrire, questionnaires de recrutement et d'actualisation, questionnaires d'enquêtes et de satisfaction, menus d'aide, invitations, réponses aux interrogations des panélistes... (16, 17). Pour animer l'*access panel*, on peut utiliser maints outils : courriels de *feedback*, *newsletters*, forum ou espace « panélistes » du site de la société d'études, ligne d'assistance aux répondants, enquêtes de satisfaction, institution d'un référent par panéliste... (6, 16, 17, 21). Sur un plan plus général, est souvent invoquée la nécessité, dans le contexte du Web 2.0, d'évoluer vers des schémas plus participatifs, plus communautaires (7).

D'autant plus que l'environnement Internet offre un large éventail de possibilités et génère sans cesse de nouvelles sources de distraction, souvent bien plus attrayantes et gratifiantes que la participation à des enquêtes en ligne, qui apparaissent trop souvent comme des réminiscences des premiers temps d'Internet (21). Ce qui impose de faire preuve de créativité en matière de stratégie de communication et de s'assurer que les panélistes sont suffisamment valorisés pour l'aide qu'ils apportent (21).

3/ Une approche loyale : les panélistes sont très sensibles à la confidentialité des données et à la transparence de l'information. Particulièrement en ce concerne l'identité et les objectifs du propriétaire de l'*access panel* et l'objet et la longueur annoncée des enquêtes (6, 17).

4/ Une attention particulière portée aux nouveaux panélistes : l'intérêt soutenu suscité par le caractère novateur de l'expérience s'étiolerait au cours des premiers mois. Les panélistes

nouvellement recrutés doivent donc être sollicités rapidement après leur inscription pour limiter la mortalité précoce (7).

5/ Une gestion intelligente des départs : on veillera à ce que les membres qui se retirent gardent une bonne impression de leur expérience de panéliste, d'une part pour promouvoir une image positive de la profession et, d'autre part parce qu'il est avéré qu'un panéliste sur le départ peut se raviser si on lui en offre l'opportunité (16).

6/ Des mesures de satisfaction : il est préconisé de combiner deux types de mesures : la satisfaction propre à chaque enquête, qui doit être consignée dans le rapport d'étude et les indices globaux de satisfaction (17).

7/ Des études suffisamment diverses : Proposer un large éventail de sujets, de types d'études et de formes de questionnaires peut avoir un effet positif sur la participation durable des sujets et leur satisfaction à long terme (7). Il convient notamment de ne pas abuser des *trackings*, susceptibles de lasser les panélistes.

8/ Un volume de sollicitations raisonnable : une sollicitation excessive, au-delà d'un éventuel effet de conditionnement, peut accroître les non réponses et l'attrition, ce qui a conduit les organismes professionnels à prescrire des limites à la fréquence ou au nombre des requêtes. Toutefois, il ne faut pas oublier qu'un manque de stimulation peut également être préjudiciable en amenant le panéliste à perdre de vue son appartenance à l'*access panel*, voire à le quitter (16, 17).

Les spécialistes ne s'accordent toutefois pas sur les seuils critiques. Le nombre idéal de sollicitations dépend il est vrai de multiples facteurs : motivations des panélistes, qualité des relations, politique de récompenses, intérêt et diversité des sujets, longueur des questionnaires...

De plus, le débat sur le nombre de sollicitations optimal est faussé par la multi-inscription, qui rend impossible le contrôle du volume exact de sollicitations que peut recevoir un panéliste (22).

La gestion des enquêtes

Les études menées à partir des *access panels* peuvent générer deux types d'erreur affectant la représentativité de l'échantillon final : une erreur de sélection et une erreur de non réponse.

L'erreur d'échantillonnage ne peut ici être calculée, l'*access panel* n'étant pas constitué, sauf exception, sur une base probabiliste.

L'erreur de sélection

Son importance dépend du nombre de panélistes sélectionnés et de la différence entre les panélistes qui sont sélectionnés et ceux qui ne le sont pas sur les variables étudiées. Le respect de règles strictes de sélection est un impératif. En effet, il peut être tentant de générer facticement un taux de retour satisfaisant en sollicitant les panélistes les plus coopératifs ou les plus prompts à répondre, avec les biais qui peuvent en résulter (2). La règle de base est de ne procéder que sur invitation en s'assurant que chaque individu sélectionnable a une probabilité égale ou du moins connue d'être contacté. On peut opérer un tirage aléatoire simple ou stratifier l'échantillon en fonction de critères sociodémographiques, de critères relatifs à l'historique des répondants (ancienneté, nombre d'invitations adressées) ou au profil d'internaute (fréquence et durée de l'utilisation d'Internet) (21, 22). On peut également surpondérer les strates correspondant aux segments dont le taux de réponse attendu est le plus bas afin de limiter la nécessité de pondérer *a posteriori* (16). L'utilisation des quotas est fréquente, non seulement pour éviter les redressements sociodémographiques, mais aussi pour ne pas avoir à récompenser des répondants surnuméraires dans certaines catégories (les cellules se ferment dès lors que le nombre requis de participants est atteint). Cette méthode pose de sérieux problèmes, notamment parce qu'elle favorise les internautes les plus réactifs, qui sont souvent des multi-inscrits et des utilisateurs intensifs d'Internet (22).

Par ailleurs, les comportements des panélistes peuvent générer une erreur additionnelle. Ainsi, si certains individus déclarent à tort être qualifiés pour une enquête, apparaît une erreur de sur-couverture par rapport à la population mère de l'enquête en question (2). Une parade, utilisée notamment par *Knowledge Networks*, consiste à remettre des récompenses à toutes les personnes sollicitées, qu'elles soient ou non qualifiées.

L'erreur de non réponse

Son importance dépend de deux facteurs : la proportion de répondants et la différence entre répondants et non répondants sur les variables étudiées. Le premier facteur est souvent celui qui est jugé prioritaire, d'une part parce que c'est celui qui préoccupe le plus les instituts (la participation aux enquêtes est en net déclin et des taux inférieurs à 20% ne seraient plus aujourd'hui inhabituels (19)) et, d'autre part, parce que c'est celui sur lequel il est le plus aisé d'agir. En réalité, il importe d'essayer d'agir sur les deux facteurs, plusieurs études ayant

montré qu'un taux de réponse élevé n'entraînait pas nécessairement une amélioration des résultats de l'étude (7, 22).

Il faut être conscient par ailleurs que l'on ne peut comparer le taux de réponse à une enquête auprès de volontaires avec celui d'une enquête probabiliste, puisqu'il conviendrait en toute logique de prendre aussi en compte le taux de refus durant la phase de recrutement (2, 12). La notion de non réponse recouvre en outre différentes réalités et peut-être qualifiée de méta information (12). Tout d'abord, certains internautes ne peuvent être contactés ou se trouvent dans l'impossibilité de répondre alors qu'ils étaient disposés à le faire (changement d'adresse, filtres anti-spam, incapacité à accéder au questionnaire ou à certains éléments pour des raisons techniques, prise de connaissance trop tardive de l'invitation, impossibilité de respecter la date limite). Cette impossibilité de répondre serait d'ailleurs le principal motif de non réponse selon Comley (6). Il est également fréquent qu'une partie des internautes sollicités ne soient pas qualifiés pour l'enquête. Enfin, le taux d'abandon en cours de questionnaire peut être très élevé et atteindre jusqu'à 40% (5). De fait, pour cerner la non-réponse aux différents stades et mieux définir les mesures correctrices idoines, il convient de calculer les indicateurs suivants (encadré 4), comme le préconisent d'ailleurs les organismes professionnels.

Encadré 4 – Les indicateurs de non réponse

Nombre de courriels d'invitation adressés

Proportion d'invités ayant ouvert le questionnaire

Proportion d'invités ayant répondu aux questions de filtrage

Proportion d'invités ayant abandonné en cours

Proportion d'invités ayant complété le questionnaire

Proportion d'invités dont le questionnaire est validé

Les recherches montrent que certaines caractéristiques du panéliste et de l'enquête peuvent avoir un impact sur le taux de réponse et l'erreur de non réponse. Nous ne traiterons pas ici de nouveau des facteurs qui peuvent jouer sur l'attrition et créent un contexte plus ou moins favorable à un taux de réponses élevé pour une enquête donnée. Il va de soi par ailleurs que l'absence de purge des panélistes inactifs réduit *ipso facto* le taux de réponse moyen.

- ***Les caractéristiques de l'enquête***

De nombreux facteurs ont fait l'objet d'études auprès de panélistes.

1/ La durée de l'enquête : la réactivité des panélistes n'est pas homogène. Ainsi, Jourdan et Jourdan (12) observent que 50% des répondants complètent le questionnaire dans les 24 heures et à peu près 70% dans les deux jours tandis que Vonk, van Ossenbruggen et Willems (22) aboutissent à un taux de retour de 75% en 24 heures et de 97% en quatre jours. Une distorsion de l'échantillon peut en résulter. Seraient plus réactifs en particulier les internautes qui passent beaucoup de temps sur Internet et ceux qui bénéficient du haut débit (12). Il y aurait également une corrélation entre le temps de réponse à une enquête, le délai de réaction lors des enquêtes antérieures et le délai d'enregistrement initial, ce qui suggère une certaine constance dans le comportement du panéliste (12). D'une manière générale, comme un délai trop court pénalise les panélistes qui consultent rarement leur messagerie, la durée de l'étude doit être d'autant plus longue que le thème est lié au profil d'internaute (*e. g.* achats en ligne).

2/ Le thème de l'enquête : non seulement les internautes intéressés par le sujet seraient plus enclins à répondre ce qui se traduirait notamment par une surestimation des intentions d'achat lors des tests de concept, mais il apparaît que l'intérêt pour les thèmes d'étude est fort variable. Comley (6) relève ainsi des scores d'intérêt supérieurs à 60% pour le magasinage, la nourriture, les produits pour la maison et les équipements de loisirs, la politique (33%) et le sport (30%) étant en queue de peloton. Il en résulte que la motivation des panélistes à participer à des enquêtes en général et à donner leur opinion est essentielle.

3/ Le questionnaire : participer à une enquête doit être une expérience agréable pour le panéliste et un questionnaire mal conçu ou trop austère peut affecter le taux de réponse (1, 5, 10). En particulier, certaines questions semblent fort peu prisées des panélistes (questions numériques, questions ouvertes et grilles combinant plusieurs questions) (6). Les panélistes avancent par ailleurs volontiers la longueur comme motif de non participation ou d'abandon. Lors de l'étude de Comley (6), 35% des panélistes situent le seuil critique à 15 minutes et 82% à 30 minutes, tandis que Namiranian, Moskowitz et Gollman (14) relèvent que c'est entre 16 et 20 minutes que la majorité des panélistes expriment un début de lassitude.

4/ L'invitation : comme nous l'avons souligné, l'étude de Namiranian, Moskowitz et Gofman (14) a mis en évidence les arguments ayant le meilleur impact sur les quatre segments de panélistes identifiés en termes de motivations. Ils estiment qu'un message approprié pourrait augmenter le taux de participation de 20%.

5/ Les récompenses : l'éventail des possibilités est très large : argent, points convertibles en argent, en chèques-cadeaux ou en cadeaux, cadeaux, produits testés, réductions de prix, participation à des loteries, dons à des œuvres de charité, rapports d'études... L'argent prédomine, suivi des loteries, mais quelques sociétés, telle IMRS au Royaume-Uni, opèrent avec succès sans pour autant recourir à des récompenses (6). Idéalement, les récompenses doivent être suffisamment stimulantes pour inciter les panélistes à s'inscrire et à répondre, tout en restant assez modiques pour ne pas constituer une véritable source de revenus. Les études sur les attentes des panélistes en matière de politique de récompense, placent systématiquement en tête les récompenses monétaires, loin devant les autres formes de récompenses matérielles et très loin devant les bénéfices non matériels (rapports d'étude, dons à des œuvres de charité...) (6, 10, 16). Par ailleurs, il ressort des études conduites que les récompenses certaines ont plus d'effet que les récompenses hypothétiques (loteries) et des récompenses prépayées lors de l'invitation sont plus efficaces (mais plus coûteuses) que des récompenses promises. En tout état de cause, puisque les différentes formes de récompenses peuvent exercer un attrait variable sur les segments de panélistes, il semble judicieux de les diversifier pour avoir des *access panels* mieux équilibrés, limiter l'erreur de non-réponse et accroître la satisfaction des panélistes (6, 8, 17). On peut également offrir des choix aux panélistes, en leur proposant par exemple de reverser leurs gains à une œuvre de charité (10).

6/ Les relances : elles sont un moyen de diminuer le taux de non-réponse bien moins onéreux que pour les autres modes d'enquête. Jourdan et Jourdan (12) constatent que 21% des répondants à leur étude ont participé suite à une relance. Les relances seraient d'autant plus utiles que le panéliste manifeste globalement de l'inertie (délai d'enregistrement, temps moyen pour répondre aux enquêtes précédentes...).

- *Les caractéristiques du panéliste*

La relation entre le profil du panéliste et sa propension à participer a été notamment étudiée par Jourdan et Jourdan (12) et Vonk, van Ossenbruggen et Willems (22). Jourdan et Jourdan concluent de leur étude que la propension à participer dépend avant tout du profil de l'individu en tant que panéliste (nombre d'enquêtes complétées, délai moyen pour répondre, ancienneté, délai initial pour s'enregistrer), puis de son profil sociodémographique (activité, âge, revenu, genre) et enfin de son profil d'internaute (volume de connexion à Internet depuis son domicile). Il apparaît que le taux de réponse tend à décroître avec l'ancienneté de la présence dans l'*access panel* (12, 22), que les multi-inscrits ont une propension plus forte à

répondre et que le mode de recrutement a un impact significatif, le taux de réponse étant plus élevé chez les panélistes qui ont été recrutés de manière active plutôt que par autosélection (22). Sur le plan sociodémographique, les résultats des études disponibles sont très contradictoires. Nous noterons simplement que l'étude la plus ample (22) laisse apparaître des taux de réponse sensiblement supérieurs chez les seniors et les femmes

Les procédures de redressement

Le cumul des biais peut générer une distorsion importante de l'échantillon qui aura répondu *in fine* à une enquête particulière et affecter gravement la validité des données recueillies. Des écarts considérables ont parfois été relevés entre les réponses apportées par les panélistes en ligne et des échantillons interrogés par d'autres voies ou des données statistiques. Les panélistes en ligne auraient ainsi une propension plus marquée à adopter les innovations, en particulier dans les nouvelles technologies (19). Des écarts ont été observés sur d'autres thèmes comme les comportements électoraux, la possession de produits financiers, la fréquentation des médias ou les comportements d'achat et de consommation. Les comparaisons se heurtent toutefois au problème de l'équivalence des réponses, car une partie de l'écart peut être imputé, non pas à des problèmes de représentativité, mais aux effets du mode d'enquête (3). Par rapport aux enquêtes téléphoniques, les enquêtes en ligne seraient ainsi caractérisées par de moindres biais d'acquiescement et de désirabilité sociale, par des réponses plus nuancées, une moindre spontanéité, des scores de reconnaissance plus élevés et une plus grande richesse de l'information (1, 3, 15). Cet effet du mode est d'ailleurs le principal obstacle, avec le coût, à l'utilisation d'approches multimodales.

Les instituts ont recours à des procédures de redressement pour tenter de remédier aux biais. Ce sont les variables sociodémographiques qui sont le plus souvent utilisées pour pondérer les données d'enquête. Toutefois, cette procédure est basée sur la croyance que la coïncidence entre deux échantillons sur un éventail de caractéristiques démographiques va l'assurer également sur les variables d'intérêt, ou en d'autres termes, « la présomption sous-jacente est que les femmes âgées vivant seules, les personnes à faible niveau d'études, les minorités ethniques ou les autres groupes habituellement sous-représentés qui ne participent pas à l'étude ont les mêmes caractéristiques que ceux qui y participent. Dans certains cas, cela peut être une présomption plausible ; dans d'autres, ce ne l'est certainement pas, et, dans la plupart des cas, il est difficile de le tester » (Bethlehem et Stoop, 2007) (2). Le risque d'avoir des

répondants atypiques est particulièrement élevé dans les groupes où la pénétration d'Internet est faible. Il a ainsi été observé que les répondants seniors en ligne avaient des schémas de consommation distincts des seniors en général, phénomène qui a été attribué à une différence d'âge cognitif (15).

Une méthode de redressement plus élaborée consiste à procéder à une pondération par les scores de propension, afin de capturer la probabilité de répondre. Concrètement, il s'agit de mesurer des variables qualifiées de psychographiques ou de webographiques et qui sont supposées refléter les différences entre la population totale et les répondants en ligne. Les mêmes questions sont posées lors des enquêtes en ligne à calibrer et au cours d'études de référence menées par RDD. Des scores de propension sont obtenus à partir d'une régression logistique sur les variables webographiques. Les deux populations de répondants sont ensuite stratifiées en fonction des scores et on procède à une pondération pour faire coïncider les proportions de répondants dans chaque strate pour l'enquête en ligne et l'enquête de référence. Cette méthode présente plusieurs limites. Tout d'abord, déterminer les variables qui capturent le mieux la propension à répondre lors d'une enquête spécifique n'est guère évident. A titre d'exemple, Harris Interactive s'appuie sur plusieurs séries de questions portant sur des attitudes (sentiment que sa vie privée est violée par certaines pratiques comme les cookies, le télémarketing, les contrôles aux aéroports ; curiosité à l'égard des nouveautés...), et des faits (connaissance d'au moins une personne à orientation sexuelle particulière, lecture d'un livre dans le mois écoulé...) (20). La réalisation d'une étude de référence augmente également le coût des enquêtes, même si elle sert à caler plusieurs enquêtes et elle peut également souffrir d'un biais de non réponse non négligeable, les enquêtes téléphoniques étant de moins en moins représentatives de l'ensemble de la population (2). Enfin, le calage des réponses peut être faussé par l'effet du mode d'enquête.

La plupart des études aboutissent à la conclusion que les pondérations sociodémographiques ou webographiques n'aboutissent qu'à des améliorations partielles, parfois marginales (20). Pour limiter la nécessité de pondérations à l'issue de l'enquête, on peut procéder à une stratification non proportionnelle *a priori*, en modulant le nombre d'invitations adressées dans chaque strate en fonction de l'historique des réponses. Par exemple, si les résultats antérieurs laissent apparaître un taux moyen de réponses de 40% chez les hommes et de 60% chez les femmes, l'échantillon sélectionné intégrera ces proportions afin de mieux équilibrer l'échantillon final. Cette méthode limite l'ampleur des pondérations *a posteriori*, ce qui évite

d'appliquer des poids déraisonnables à de petits groupes de répondants appartenant à des strates très nettement sous-représentées. Mais, idéalement, il convient de limiter la nécessité de recourir à des redressements en prenant les mesures permettant de pallier au maximum les biais à tous les stades de la constitution et de la gestion d'un *access* panel et des échantillons sondés pour chaque enquête particulière.

Conclusion

Si le monde des *access* panels en ligne a mené dans les premiers temps une course à la puissance, la prise de conscience de l'exigence de qualité est très nette depuis quelques années. Il faut d'ailleurs se garder d'idéaliser les méthodes traditionnelles. Comme le soulignent Schillewaert et Meulemeester (18), « même s'il fait l'objet de moins de critiques, le problème de la pénétration incomplète est également pertinent pour l'enquête téléphonique. Avec l'apparition de la téléphonie mobile, de nombreuses personnes n'ont plus de ligne fixe à domicile, ce qui rend les procédures RDD inadéquates pour des échantillonnages probabilistes des populations nationales. De plus, toutes les méthodes d'enquête pâtissent dans une certaine mesure du biais d'auto-sélection. En conséquence des échantillons aléatoires purs sont très difficiles à obtenir et les chercheurs doivent toujours être prudents quant au caractère généralisable des résultats ».

S'il convient donc de relativiser les difficultés méthodologiques que présentent les *access* panels en ligne, il ne s'agit pas pour autant de les minorer. D'autant plus qu'il est vraisemblable que les *access* panels qui se prêtent à des recherches sont parmi les plus satisfaisants du marché et que les résultats obtenus tendent à sous-estimer les problèmes de représentativité effectivement rencontrés. Sauf à imaginer qu'on puisse transformer le plomb en or, c'est bien à notre sens dès l'origine – c'est-à-dire lors de la constitution de l'*access* panel que des mesures doivent être prises pour réduire les sources d'erreur. Comme le souligne Harlow (11), « il n'y a pas d'alternative à la conception et à la construction de panels avec un recrutement et un management minutieux dès l'origine. Cela génère des coûts, notamment financiers. Concevoir des panels soigneusement recrutés par téléphone est plus coûteux que de les générer par des invitations sous formes de *pop-ups*, de listes d'adresses *e-mail* achetées ou d'autres formes de recrutement en ligne, et notre expérience nous indique que nombre de clients ne sont pas prêts à s'engager pour un panel soigneusement recruté, alors qu'ils peuvent obtenir deux fois plus de répondants pour le même prix auprès d'un panel

non probabiliste. Nous pensons toutefois que, au fur et à mesure que les clients gagneront en expérience, il s'en trouvera davantage à opter pour des échantillons soigneusement constitués et avec vérification de l'identité des répondants ». Nous pouvons rajouter que la réduction initiale de l'erreur de couverture et de l'erreur d'auto-sélection a, par effet de cascade, un impact positif sur l'ensemble de la chaîne de qualité d'une enquête : taux de réponse plus élevé, taux d'attrition moindre, erreur de mesure réduite. Plus précisément, nous considérons que quatre mesures doivent être prioritairement préconisées :

- 1/ Construire l'échantillon à partir de sollicitations hors ligne
- 2/ Ne pas se limiter aux critères sociodémographiques pour assurer la représentativité, mais également aux critères en relation avec Internet (fréquence de l'utilisation, ancienneté de l'utilisation...).
- 3/ Empêcher toute forme d'auto-inscription
- 4/ Ne pas mettre excessivement l'accent sur les récompenses monétaires ou matérielles et stimuler les motivations intrinsèques.

Par ailleurs, si de nombreuses pistes de recherches sont envisageables, trois priorités nous paraissent devoir être avancées :

1/ Les recherches conduites sont, sauf rares exceptions, des sondages menés en ligne auprès des panélistes actifs. Il pourrait être instructif pour mieux orienter les règles de recrutement et de management des *access* panels, de mener des consultations par d'autres voies (téléphone, courrier, face-à-face), de recourir à des études qualitatives et d'étudier les autres publics susceptibles de nous apporter un éclairage : internautes ayant décliné l'invitation à s'inscrire, panélistes ayant décidé de se désinscrire, panélistes n'ayant pas participé à une enquête ou ayant abandonné en cours. De telles études pourraient nous permettre de recueillir des informations plus riches et plus spontanées sur les motivations et les freins à s'inscrire, à participer aux enquêtes et à compléter les questionnaires et, *in fine*, de mieux orienter les règles de recrutement et de management permettant d'améliorer la représentativité des *access* panels et des échantillons qui en sont issus.

2/ Les alternatives au tout électronique dans le recrutement et le management des *access* panels sont à considérer avec beaucoup d'attention. En particulier, l'influence du facteur humain à différents stades mérite d'être évaluée, même si l'incidence sur les coûts ne peut

évidemment être ignorée. Les études de Körner et Nimmergut A. (13) et de Harlow (11) montrent que recruter hors ligne en s'appuyant sur des contacts personnels est susceptible de générer des taux d'inscription sensiblement supérieurs et de permettre l'élimination des fraudeurs. La relation humaine semble également de nature à avoir des effets positifs sur le management de l'*access panel* (17). Tout particulièrement, la mise en place d'un dispositif de référents pour les panélistes paraît de nature à renforcer les liens et l'attachement à l'institut d'études, réduire l'attrition et les non-réponses, consolider les motivations intrinsèques et, in fine, améliorer la qualité des données d'enquête. Ces référents pourraient utiliser la voie électronique, mais également d'autres modes de contact (téléphone, voire face-à-face) pour remotiver les panélistes qui ne répondent pas aux enquêtes, les relancer ou les consulter en cas d'insatisfaction. On peut à cet égard penser que la réussite de l'*access panel* de *Knowledge Networks* s'explique non seulement par la manière dont est traité le problème de couverture, mais aussi par la place accordée au contact humain : le recrutement se fait par téléphone, les panélistes sont systématiquement appelés s'ils n'ont pas répondu à deux enquêtes consécutives, une ligne d'assistance gratuite avec du personnel est ouverte de neuf heures du matin à neuf heures du soir. Les courriers sont également utilisés notamment pour la prénotification des enquêtes.

3/ Il serait pertinent, en s'appuyant notamment sur les travaux de Comley (6) et Sassinot-Uny et Gadeib (17), d'approfondir l'analyse des déterminants de la satisfaction des panélistes. Non seulement, parce que la satisfaction réduit l'attrition et le taux de non réponses, mais aussi parce que les panélistes ne sont pas une ressource illimitée (7, 17, 21). Il devient donc de plus en plus urgent d'identifier avec précision les facteurs qui contribuent à ce que chaque enquête soit une expérience aussi gratifiante que possible et les outils relationnels et communautaires qui participent à la fidélisation des panélistes dans un contexte qui n'est plus celui de l'émergence d'Internet.

Références bibliographiques

- 1/ Baker R. (2008), Without a net – Making business decisions from online panel data, *Panel Research 2008, Esomar World Research Conference*, Dublin.
- 2/ Bethlehem J. et Stoop I. (2007), Online panels – A paradigm theft, in *The challenges of a changing world*, M. Trotman et al. eds, Association for Survey Computing, 113 – 131.

- 3/ Bronner F. et Kuijlen T., The live or digital interviewer – A comparison between Capi, Capi and Cati with respect to differences in response behaviour, *International Journal of Market Research*, 49, 2, 167 – 190.
- 4/ Burn M. et Thomas J. (2008), Plus vite et moins cher : avons-nous encore besoin de bonnes études de marché ?, *Revue française du marketing*, 217, 2/5, 9-26.
- 5/ Coen T., Lorch J. et Piekarski L. (2005), The effects of survey frequency on panelists' responses, *Esomar Panel Research Conference*, Budapest.
- 6/ Comley P. (2005), Understanding the online panelist, *Esomar Panel Research Conference*, Budapest.
- 7/ Comley P. (2007), Online market research, in *Market research handbook*, 5ème édition, Esomar World Research Publication, M. van Hamersveld et C. de Bont eds, John Wiley & Sons, 401-419.
- 8/ Göritz A. S. (2004), The impact of material incentives on response quantity, response quality, sample composition, survey outcome, and cost in online access panels, *International Journal of Market Research*, 46, 3, 327-344.
- 9/ Göritz A. S. (2004), Recruitment for online access panels, *International Journal of Market Research*, 46, 4, 411-425.
- 10/ Guidi M. (2007), Understanding respondent motivation, SSI White Paper n° 9, Survey Sampling International, www.surveysampling.com.
- 11/ Harlow B. (2008), Analytic implications of panel data quality – The role of panel recruitment, *Panel Research 2008, Esomar World Research Conference*, Dublin
- 12/ Jourdan P. et Jourdan V. (2005), An exploratory study of the participation in online surveys using access panels, *Esomar Panel Research Conference*, Budapest.
- 13/ Körner T. et Nimmergut A. (2004), Using an access panel as a sampling frame for voluntary household surveys – Experiences from a pilot study, *Statistical Journal of the United Nations*, Economic Commission for Europe, IOS Press, 21, 33 – 52.
- 14/ Namiranian L., Moskowitz H. R. et Gofman A., Motivations to join - What language do online panellists need to hear to volunteer? *Panel Research 2006, Esomar World Research Conference*, Barcelone.
- 15/ Nauckhoff F., Östgren F. et Mattsson B. (2007), Moving tracking studies on-line - Experiences gained when using fully automated procedures, *Panel Research 2007, Esomar World Research Conference*, Orlando.
- 16/ Postoaca A. (2006), *The anonymous elect – Market research through online access panels*, Berlin-Heidelberg, Springer.

- 17/ Sassinot-Uny L. et Gadeib A. (2007), Panel satisfaction index – Quality target for online access panels owners? *Panel Research 2007, Esomar World Research Conference*, Orlando.
- 18/ Schillewaert N. et Meulemeester P. (2005), Comparing response distributions of offline and online data collection methods, *International Journal of Market Research*, 47, 2, 163-178.
- 19/ Schillewaert N., Verhaeghe A., Weijters B. et de Wolf K. (2006), Social class and life style differences between modes of data collection, *Panel Research 2006, Esomar World Research Conference*, Barcelone.
- 20/ Schonlau M, van Soest A. et Kapteyn A. (2007), Are webographic or attitudinal questions useful for adjusting estimates from Web surveys using propensity scoring?, *Survey Research Methods*, 1, 3, 155-163.
- 21/ Van Walwyk M. et Garland C. (2008), Turning the super tanker – The migration from a postal to online methodology, *Panel Research 2008, Esomar World Research Conference*, Dublin
- 22/ Vonk T., van Ossenbruggen R. et Willems P. (2006), The effects of panel recruitment and management on research results, a study among 19 online panels, *Panel Research 2006, Esomar World Research Conference*, Barcelone.

¹ Access panel est parfois traduit par « panel ouvert », mais cette dénomination reste assez peu répandue.

² A été notamment créée une base de données des répondants à problème : *PureSample.com database*. Cette liste noire alimentée par les utilisateurs et fournisseurs d'access panels comprend plus de 1,3 million d'adresses e-mail.

³ Une multitude de sites sont clairement dédiés aux chasseurs de récompenses (e. g. www.sondages-remuneres.com, www.anosleskados.com/sondages-remuneres.htm, www.argent-clic.com/avis-remuneres.htm, www.sondage-avenue.com/sondage.htm)