



# DIRECTIVE ESOMAR/GRBN SUR LA QUALITÉ DES ÉCHANTILLONS EN LIGNE

ESOMAR, association mondiale pour les études de marché, études sociales et d'opinion, est l'organisation essentielle pour encourager, faire avancer et élever le secteur des études de marché. [www.esomar.org](http://www.esomar.org)

GRBN (Global Research Business Network), réseau international pour la recherche professionnelle, regroupe 38 associations de recherche et plus de 3 500 sociétés spécialisées dans la recherche issues des cinq continents. [www.grbn.org](http://www.grbn.org)

© 2015 ESOMAR et GRBN. Publication : janvier 2015.

Cette directive a été rédigée en anglais et le texte anglais constitue sa version définitive. Le texte peut être copié, distribué et transmis sous réserve de mentionner les auteurs de manière appropriée et d'inclure l'avis suivant : "© 2015 ESOMAR et GRBN".

[Official Translation Partner:](#)  
[Language Connect](#)



# TABLE DES MATIÈRES

<b>1</b>	<b>INTRODUCTION ET PORTÉE</b>	<b>4</b>
<b>2</b>	<b>DÉFINITIONS</b>	<b>5</b>
<b>3</b>	<b>EXIGENCES CLÉS</b>	<b>8</b>
3.1	L'identité prétendue de chaque participant au projet de recherche doit être vérifiée.	8
3.2	Les fournisseurs doivent s'assurer qu'aucun participant aux projets de recherche ne réponde à la même enquête plus d'une fois.	9
3.3	L'engagement des participants aux projets de recherche doit être calculé et doit faire l'objet d'un rapport	11
3.4	L'identité et les données personnelles des participants aux projets de recherche doivent être protégés.	12
3.5	Soyez particulièrement vigilants avec les enfants et les jeunes personnes	12
3.6	Avoir recours à de bonnes pratiques de création de questionnaire en ligne	13
3.7	Les fournisseurs d'échantillons en ligne doivent être transparents avec les chercheurs et les clients au sujet des sources de leurs échantillons, du processus d'échantillonnage et de ses résultats.	13
3.8	Les chercheurs utilisant des échantillons en ligne doivent être transparents avec leurs clients.	14
3.9	Collecte passive de données	15
3.10	Se conformer à l'ensemble des lois, réglementations et codes de conduite du secteur pertinents.	15
<b>4</b>	<b>CONSIDÉRATIONS SPÉCIALES</b>	<b>15</b>
4.1	L'exclusion par catégorie et autres types d'exclusion	15
4.2	Mobile	16
4.3	Automatisation des enquêtes et de l'échantillonnage	17
<b>5</b>	<b>L'ÉQUIPE DU PROJET</b>	<b>17</b>
<b>6</b>	<b>RÉFÉRENCES</b>	<b>17</b>

# 1 INTRODUCTION ET PORTÉE

Avec l'émergence et l'acceptation générale des échantillons en ligne pour les études de marché, études sociales et d'opinion apparaît la responsabilité de mesurer et de s'assurer de la qualité des résultats des recherches faisant appel à de tels échantillons. Un certain nombre de préoccupations concernant la qualité des échantillons ont été soulevées dans tout le secteur. Avec notamment des préoccupations concernant :

- Des participants aux projets de recherche professionnels qui tentent de diverses manières de maximiser leurs opportunités de participation aux enquêtes
- Des participants aux projets de recherche inattentifs voire délibérément mensongers
- La possibilité pour les personnes de dupliquer leurs participations avec les fournisseurs d'échantillons en ligne qui élargissent leurs sources en quête d'échantillons plus divers et de populations à faible incidence
- La représentativité, c'est-à-dire la mesure dans laquelle les résultats des recherches reflètent précisément la population ciblée, de quelque manière qu'elle ait été définie

La présente directive répond à ces préoccupations en présentant des conseils quant aux exigences opérationnelles concernant la provision d'échantillons en ligne pour les études de marché, études sociales et d'opinion. Elle expose des méthodes à utiliser par les fournisseurs d'échantillons en ligne, les acheteurs et les clients finaux, afin de garantir que l'échantillon réponde aux critères de qualité largement admis. Sa lecture est recommandée à toutes les parties prenantes du processus de recherche, des concepteurs de l'enquête aux utilisateurs de données.

Elle est vouée à s'appliquer à tous les types d'échantillons en ligne, y compris ceux recrutés auprès de panels, de réseaux sociaux, et par des méthodes d'interception web. Elle n'est pas conçue pour les échantillons fournis par les clients, tels que les listes de clients. Elle n'est pas non plus conçue pour les échantillons recrutés pour les projets de recherche qualitatifs, tels que les groupes de discussion ou les entretiens individuels en face à face, puisque les interactions directes entre l'enquêteur et le participant au projet de recherche constituent des opportunités d'assurance qualité qui sont généralement inexistantes dans les projets de recherche quantitatifs auto-administrés. Néanmoins, les enquêteurs pourraient trouver utiles certaines des pratiques suggérées lorsqu'ils travaillent avec ces sources d'échantillons.

Cette directive n'a pas pour but de se substituer à la lecture attentive et à la compréhension du [Code international ICC/ESOMAR des études de marché et d'opinion](#), qui a été adopté par plus de 60 associations locales dans le monde entier, ni aux codes individuels des 38 associations qui constituent le [GRBN](#). Elle a au contraire été conçue comme une interprétation des principes fondateurs de ces codes, dans le contexte de la recherche en ligne. Elle repose également sur un certain nombre de sources concernant ses principes de base, y compris les normes ISO 20252 – Études de marché, études sociales et d'opinion et ISO 26362 – « Access panels » pour les études de marché, études sociales et d'opinion. Elle recommande une transparence totale avec les clients et encourage l'utilisation de termes et de définitions communes. Elle propose un ensemble de pratiques conseillées pour la qualité des échantillons en ligne, bien qu'elle ne soit pas conçue pour se substituer aux normes ISO 20252 ou ISO 26362, qui nécessitent toutes deux un audit externe d'un fournisseur d'échantillons pour contrôler la conformité. Un grand nombre d'acheteurs d'échantillons en ligne souhaitent limiter leurs achats d'échantillons aux fournisseurs détenteurs d'une certification ISO en raison des garanties supplémentaires qu'apporte un audit externe.

La présente directive ne prétend pas non plus couvrir tous les facteurs susceptibles d'avoir un impact sur la qualité de la recherche en ligne. La recherche sur la recherche dans ce domaine continue d'évoluer et il existe de nombreux problèmes non résolus. Toutefois, il

existe un ensemble limité de problèmes qui figurent à l'avant-plan malgré quelques controverses continues à propos de leur impact sur la qualité des données. Il s'agit de :

- La validation des participants aux projets de recherche
- La prévention de la fraude durant les enquêtes
- L'engagement dans l'enquête
- L'exclusion par catégorie et autres types d'exclusion
- L'échantillonnage (y compris la sélection de l'échantillon, le panachage des échantillons, la pondération, les routeurs d'enquête, la détermination des profils et le filtrage).

Tout au long de ce document, le mot « doit/doivent » est utilisé pour signifier des exigences obligatoires. Nous utilisons le mot « doit/doivent » pour décrire un principe ou une pratique que les chercheurs sont tenus de respecter afin de se conformer aux codes et aux normes existants, tels que ceux cités ci-dessus. Le mot « devrait/devraient » est utilisé pour décrire l'application et l'utilisation, lorsque les chercheurs ont la possibilité de décider d'appliquer un principe ou une pratique de diverses manières en fonction de la conception de leur projet de recherche.

## 2 DÉFINITIONS

**Access panel** Une base de données de participants potentiels à un projet de recherche qui s'engagent à coopérer lors de futures demandes de collecte de données s'ils sont sélectionnés.

**Taux d'achèvement** Le nombre de participants au projet de recherche ayant entièrement répondu à l'enquête, divisé par le nombre de participants au projet de recherche ayant commencé l'enquête ou ayant au moins visualisé la première page ou la page d'accueil.

**Consentement** L'accord informé et donné librement par une personne quant à la collecte et au traitement de ses données personnelles. Dans le cadre des études de marché, études sociales et d'opinion, ce consentement doit être donné une fois que le participant a été clairement informé de la nature des données recueillies, de la fin à laquelle ces données seront utilisées et de l'identité de la personne ou de l'organisation qui détiendra les données personnelles. Le participant au projet de recherche a la possibilité de retirer son consentement à tout moment. Une archive du consentement et de la façon dont il a été obtenu doit être conservée.

**Cookies** Les cookies sont des fichiers textes contenant de petites quantités d'information, qui sont téléchargés sur un ordinateur, un appareil mobile ou tout autre appareil, lorsqu'un utilisateur visite un site Internet. Les cookies sont ensuite renvoyés au site web d'origine à chaque visite ultérieure, ou vers un autre site web reconnaissant ce cookie.

Les cookies sont utiles dans la mesure où ils permettent à un site web de reconnaître l'appareil d'un utilisateur et de rendre la navigation sur un site web plus efficace, en enregistrant les préférences de l'utilisateur, et en améliorant de manière générale l'expérience utilisateur. Ils peuvent également permettre de garantir que les offres proposées à un utilisateur en ligne lui soient plus pertinentes et correspondent mieux à ses intérêts.

**Déduplication** Dans le cas des access panels, il s'agit d'un processus visant à éliminer les individus qui se sont inscrits plus d'une fois au même access panel afin qu'ils ne participent qu'une fois.

Dans le cadre des échantillons de recherche, il s'agit d'un processus permettant de supprimer les individus qui complètent ou tentent de compléter la même enquête plus d'une fois, que ce soit parce qu'ils ont reçu plusieurs invitations ou délibérément dans le cas d'une tentative de fraude.

**Identifiant de l'appareil** Un système technologique qui établit un ensemble de données de configuration à propos de l'appareil du participant à un projet de recherche (ordinateur, smartphone, etc.) qui peut être utilisé pour créer une empreinte pour une machine ou un appareil. De tels systèmes considèrent que l'« empreinte machine » n'identifie qu'un appareil utilisant des paramètres et des caractéristiques associés à un appareil précis ou, potentiellement, au compte utilisateur d'un individu. Les systèmes d'identifiant d'appareil s'appliquent aux ordinateurs, aux appareils mobiles et aux autres appareils susceptibles d'être utilisés pour accéder à Internet et ainsi répondre à des enquêtes.

Remarque : un identifiant d'appareil peut parfois également être appelé « empreinte numérique », « empreinte machine » ou « identifiant machine ».

**Duplication** Une situation lors de laquelle un participant à un projet de recherche tente de compléter ou complète une enquête donnée plus d'une fois. Cela peut notamment se produire lorsqu'un panéliste ou un participant à un projet de recherche est membre de plus d'un(e) panel ou source d'échantillon (le panel ou la source d'échantillon coïncident) et est sélectionné pour participer à une enquête qui est partagée par plusieurs sources de panels, et qu'il ne se souvient pas avoir précédemment participé à une enquête donnée.

**Réponse frauduleuse** Une situation lors de laquelle un participant à un projet de recherche répond consciemment aux questions d'une enquête de manière malhonnête. Un exemple fréquent est la pratique de donner de fausses réponses aux questions de sélection de manière à s'assurer la participation à une enquête.

**Fournisseur d'échantillons en ligne** Un fournisseur de services responsable de la provision et de la gestion d'échantillons en ligne issus de sources pertinentes, y compris des panels, des sources basées sur une interception web (y compris des sources d'échantillonnage aléatoire), des listes de diffusion, etc. Les fournisseurs d'échantillons en ligne peuvent proposer certains ou tous les services ci-dessous :

- Les access panels
- Le courtage d'échantillons (la pratique d'acheter et de revendre des échantillons)
- L'agrégation d'échantillons (la pratique de combiner ou de regrouper plusieurs sources d'échantillons)
- Les technologies de routage ou l'accès à des technologies de routage

Les fournisseurs d'échantillons en ligne sont susceptibles de proposer d'autres services de recherche ou en lien avec la recherche.

**Taux de participation** Le nombre de membres du panel ayant fourni une réponse exploitable, divisé par le nombre total d'invitations de participation personnelles initialement envoyées aux membres, dans le cas d'un access panel se basant exclusivement sur ce type d'invitations. La définition du taux de participation pour les approches « aléatoires » et les autres approches qui ne sont pas basées sur des e-mails, est plus compliquée, et aucune démarche n'a encore été définie comme bonne pratique.

Remarque : une réponse est considérée comme exploitable lorsque le participant au projet de recherche a fourni des réponses à toutes les questions imposées par l'enquête. Lorsqu'il est possible de déterminer le nombre d'invitations non livrées (par ex. renvoyées à l'expéditeur en raison d'une boîte de réception pleine, une adresse postale ou électronique incorrecte, un numéro de téléphone invalide), cela doit alors être pris en compte lors du calcul du taux de participation. Le nombre de membres du panel n'ayant pas reçu leur invitation pour cette enquête précise sera ensuite soustrait du nombre total de membres du panel invités à participer.

**Collecte de données passive** Des méthodes de recherche par lesquelles des données sont collectées sans l'utilisation traditionnelle de questions d'enquête.

**Méthodes de validation passives** Méthodes basées sur Internet utilisées pour établir les caractéristiques du participant à l'enquête et valider les panélistes. Ces méthodes peuvent comprendre le suivi des visites de sites web des participants au projet de recherche et des panélistes, les pages spécifiques qu'ils consultent, et les liens sur lesquels ils cliquent, puis l'utilisation de ces informations pour créer un profil. Ce profil est ensuite utilisé pour fournir des informations de validation pour les participants au projet de recherche et les panélistes.

**Participant au projet de recherche** Toute personne participant à un projet de recherche. Dans de nombreux cas, leurs données personnelles sont recueillies durant le projet de recherche en question, dans le cadre d'un entretien actif ou par des moyens passifs.

**Représentativité** La mesure dans laquelle un échantillon reflète la population cible étudiée. Un échantillon représentatif est un échantillon au sein duquel la répartition des caractéristiques importantes est à peu près la même que dans la population cible. La définition des « caractéristiques importantes » dépend généralement du ou des sujets de l'enquête. La représentativité peut être atteinte au travers de l'échantillonnage, d'ajustements post-enquête (pondération), ou d'une combinaison des deux.

**Échantillon aléatoire** Un échantillon ne provenant pas d'un panel, recruté par l'intermédiaire de publicités en ligne, d'offres, ou d'autres mécanismes d'invitation similaires en temps réel.

Remarque : les termes interception web, échantillonnage en temps réel et échantillon à provenance dynamique sont parfois utilisés pour décrire les échantillons aléatoires.

**Routeur** Un système logiciel en ligne qui filtre les participants entrants puis utilise les résultats de ce processus pour attribuer des participants aux projets de recherche à l'une des diverses enquêtes disponibles. Un routeur peut également être utilisé pour soumettre aux participants au projet de recherche d'autres questionnaires de sélection et enquêtes s'ils n'ont pas été jugés éligibles à l'issue d'un questionnaire de sélection ou à la fin d'une enquête. Les routeurs sont généralement configurés en *série* ou en *parallèle*.

Un routeur en *série* utilise généralement un processus d'acheminement par lequel un participant à un projet de recherche est sélectionné séquentiellement pour les études disponibles dans l'environnement d'acheminement. Lorsqu'il est éligible à une étude, le participant au projet de recherche est souvent immédiatement renvoyé vers cette enquête précise. Lorsque le participant au projet de recherche n'est pas jugé éligible durant le processus de présélection, il est réaffecté, c'est-à-dire qu'il est renvoyé vers un autre questionnaire de sélection (associé à une autre étude dans l'environnement de routage). Le processus se répète jusqu'à ce que le participant au projet de recherche réponde aux critères de qualification pour une enquête en cours se trouvant dans l'environnement de routage, ou, selon le règlement en vigueur (par ex. la durée du processus, le nombre de questions de sélection auxquelles le participant au projet de recherche a répondu en cliquant, le nombre de questionnaires de sélection qui ont été affectés au participant), le processus cesse et le participant au projet de recherche est remercié et exclu.

Un routeur en *parallèle* utilise généralement un processus par lequel un participant à un projet de recherche est exposé à une série de questions de présélection parmi toutes les enquêtes dans l'environnement de routage ou une partie d'entre elles, simultanément, sur la même page. Une fois que le participant au projet de recherche a répondu à ces questions de présélection, il ou elle est affecté(e) à l'une des enquêtes à laquelle il est éligible. À ce stade, une présélection plus précise peut avoir lieu avant que le participant au projet de recherche ne soit considéré comme éligible à l'enquête.

**Échantillon** Une sous-catégorie de la population cible auprès de laquelle des données seront collectées.

**Cadre d'échantillonnage** Une liste d'éléments de population ou d'autres sources appropriées dont l'échantillon est issu.

**Panachage des échantillons** La pratique de combiner plusieurs sources d'échantillons hétérogènes dans le but d'obtenir un échantillon plus cohérent et plus représentatif. Cette pratique est généralement associée à des techniques d'équilibrage lors de la sélection de l'échantillon et peut se baser sur des techniques de profilage, de notation ou d'association.

**Satisficing (tout juste satisfaisant)** Un comportement de réponse à une enquête où le participant au projet de recherche ne fournit pas un plein effort cognitif lorsqu'il répond aux questions de l'enquête. De tels comportements peuvent se manifester par la sélection de réponses non substantielles, la sélection de réponses socialement admises, des réponses linéaires aux questions en matrice, et l'acquiescement, c'est-à-dire la tendance à être d'accord avec n'importe quelle assertion, quel qu'en soit le contenu.

## 3 EXIGENCES CLÉS

### 3.1 L'identité prétendue de chaque participant au projet de recherche doit être vérifiée.

Les chercheurs comme les clients ont longtemps dû faire face au problème de voir certaines personnes prétendre à de fausses identités de manière à maximiser leurs chances de participation à des enquêtes et ainsi de recevoir des récompenses ou dédommagements. En conséquence, il est devenu commun pour les fournisseurs d'échantillons en ligne dans certaines branches du secteur de vérifier l'identité prétendue de chaque participant aux projets de recherche, que ce soit au stade de l'inscription pour les panels en ligne ou au stade de l'enquête lui-même pour les sources qui ne sont pas issues d'un panel.

Une étude récente menée par l'Advertising Research Foundation (ARF) et d'autres a remis en question l'avantage de la validation d'identité pour l'amélioration de la qualité des données, avec peu ou aucune différence dans les schémas de réponse parmi les participants qui ont été validés avec succès et ceux qui n'ont pas été validés, et ceux qui refusent de fournir les informations personnelles nécessaires à la validation. En outre, cette étude a prouvé que la validation de l'identité pouvait également causer la réduction du nombre de participants disponibles en éliminant des participants qui ne sont pas en fraude. Néanmoins, un grand nombre de clients continuent d'exiger une certaine forme de contrôle des participants et la norme ISO 26362 l'exige également, dans la mesure du possible. Les chercheurs doivent également garder à l'esprit que ces résultats concernent les échantillons de consommateurs et que le contrôle reste crucial pour les échantillons B2B, en particulier dans le cas d'échantillons constitués de médecins et d'autres professionnels agréés.

Les variables spécifiques utilisées pour la validation dépendront de la méthode de contrôle utilisée, des sources disponibles pour la validation, et des restrictions, le cas échéant, imposées par les lois et réglementations en vigueur. Les variables fréquemment utilisées comprennent :

- Nom complet
- Adresse postale
- Numéro de téléphone
- Année/Date de naissance
- Adresse électronique

Si les variables ci-dessus ne sont pas disponibles, si les lois et réglementations en vigueur interdisent leur utilisation, ou si la culture et le contexte locaux s'opposent à leur utilisation, alors le fournisseur d'échantillons en ligne peut utiliser d'autres méthodes appropriées, y compris des méthodes passives (sous réserve que lesdites méthodes passives ne soient pas interdites par les lois et réglementations en vigueur).

Les sources disponibles (qui peuvent inclure des prestataires de services de validation d'adresse, des fournisseurs de services d'information, etc.) varieront également en fonction de la zone géographique et de la population étudiée. Les sources les plus appropriées sont celles qui disposent d'informations sur une très large proportion de la population cible.

La présente directive reconnaît que les échantillons ne provenant pas de panels, tels que les échantillons aléatoires et certaines formes de routage, soulèvent des problèmes importants en termes de validation, tout comme les échantillons provenant de plusieurs pays où certaines formes de validation externe sont interdites par la loi. Bien que les méthodes passives sont possibles, leur efficacité n'a pas encore été démontrée. Dans tous les cas, les sources et méthodes spécifiques utilisées, les difficultés rencontrées, et les résultats des contrôles doivent être documentés et partagés avec les clients sur demande.

Une liste étendue de variables peut s'avérer nécessaire lorsqu'il s'agit de populations spécialisées avec par exemple des médecins ou d'autres professionnels. Cette liste étendue peut inclure les données suivantes (liste non exhaustive) :

- Nom complet
- Adresse postale professionnelle
- Numéro de téléphone professionnel
- Adresse électronique professionnelle
- Numéros d'identification professionnels appropriés (si pertinent et disponible)
- Spécialité professionnelle (si pertinent et disponible)

Les sources de données utilisées pour la validation peuvent également varier en fonction de facteurs tels que le type de participant au projet de recherche ciblé ou la zone géographique ciblée. Cette directive admet que les sources de données disponibles et utiles dans un pays peuvent ne pas être disponibles et utiles dans d'autres pays. Elle reconnaît également que les sources de données utilisées pour la validation sont rarement exhaustives et que les techniques d'association automatique des informations d'identification collectées auprès des participants aux enquêtes par de telles sources peuvent résulter en de faux positifs comme de faux négatifs. Ainsi, diverses sources de données doivent être exploitées lorsqu'elles existent. En outre, étant donné ces incertitudes inévitables concernant les résultats de la validation, les fournisseurs d'échantillons en ligne sont encouragés à développer des codes de résultats qui expriment le niveau d'incertitude quant à l'identité de chaque participant au projet de recherche (plutôt que l'utilisation d'un simple indicateur binaire indiquant le succès ou l'échec). Par exemple, un participant à un projet de recherche pourrait correspondre seulement sur une sous-catégorie des données de validation collectées.

Les sources spécifiques utilisées pour la validation doivent être documentées et fournies aux clients sur demande.

Enfin, les fournisseurs d'échantillons devraient soigneusement analyser les contrôles d'identité négatifs afin de déterminer les participants au projet de recherche qui doivent être supprimés d'un panel, d'une enquête ou d'une autre activité de recherche. Les méthodes de contrôle d'identité pouvant résulter en des faux positifs comme des faux négatifs, les fournisseurs d'échantillons se doivent de faire preuve de bon sens dans cette pratique et être transparents avec leurs clients et les utilisateurs de données.

### **3.2 Les fournisseurs doivent s'assurer qu'aucun participant aux projets de recherche ne réponde à la même enquête plus d'une fois.**

Les fournisseurs d'échantillons utilisant de plus en plus des sources multiples (panels multiples, réseaux sociaux, échantillons aléatoires, etc.) pour créer leurs échantillons, il est de plus en plus probable que le(s) même(s) participant(s) au projet de recherche soit(-ent) invité(s) voire même répondent à la même enquête plus d'une fois. Les participants aux

projets de recherche dupliqués doivent être supprimés avant l'analyse, soit par le fournisseur d'échantillons, soit par le chercheur.

### **3.2.1 Identifiant d'appareil**

Une méthode de déduplication commune consiste à utiliser l'identifiant d'appareil de l'ordinateur ou de l'appareil d'un participant à un projet de recherche. Un identifiant d'appareil peut souvent également être appelé « empreinte numérique », « empreinte machine » ou « identifiant machine ». Un identifiant d'appareil est habituellement créé en utilisant des variables ou des paramètres d'un navigateur web et inclut typiquement :

- Système d'exploitation ;
- Fuseau horaire ;
- Langue ;
- Type de navigateur ;
- Paramètres du navigateur ;
- Identifiant Flash ;
- Identifiant cookie ; et
- Adresse IP.

En outre, la technologie d'identification des appareils doit être capable de supporter l'identification de géolocalisation et à la fois l'identification des doublons et des serveurs proxy dans la mesure du possible. Toutefois, les chercheurs doivent être conscients que l'adresse IP et les données de géolocalisation peuvent ne pas être complètes, puisque plusieurs ordinateurs connectés à un réseau d'entreprise ou ayant le même fournisseur de service Internet peuvent partager la même adresse IP. En outre, les participants aux projets de recherche peuvent utiliser diverses techniques pour masquer les adresses IP et les données de géolocalisation.

L'utilisation des identifiants d'appareil a soulevé des préoccupations en termes de confidentialité dans certaines juridictions. Les fournisseurs d'échantillons en ligne et les chercheurs doivent s'assurer que toute utilisation de la technologie est conforme aux lois en vigueur. La technologie d'identifiant d'appareil utilisée ne doit pas accéder à des informations personnellement identifiables et seul l'identifiant d'appareil lui-même peut être transféré ou stocké dans une base de données.

Malheureusement, la déduplication basée sur les identifiants d'appareil peut s'avérer problématique. Tout comme pour la validation de l'identité, les faux positifs comme les faux négatifs sont possibles. L'utilisation croissante des appareils mobiles pour remplir les enquêtes signifie qu'un nombre plus limité de paramètres de navigateur sont disponibles pour créer un identifiant d'appareil. Par conséquent, les fournisseurs d'échantillons en ligne et les chercheurs sont encouragés à développer des codes de résultat reflétant la certitude avec laquelle on estime que deux participants à un projet de recherche ou plus sont les mêmes (plutôt que d'utiliser un simple indicateur binaire).

### **3.2.2 Cookies et autres objets similaires**

Les fournisseurs d'échantillons en ligne utilisent ou coopèrent souvent avec des tiers qui utilisent des cookies et d'autres objets similaires, y compris des objets partagés localement (souvent appelés « cookies flash »), des balises web (y compris des gifs transparents ou clairs) pour les panels et les enquêtes. Les fins légitimes auxquelles les cookies et les autres objets similaires sont utilisés comprennent :

- L'identification de panélistes ou de participants aux projets de recherche qui sont requis pour les services demandés par l'utilisateur (pour participer aux panels et aux enquêtes) ;

- La validation et la prévention de la fraude, y compris leur utilisation légitime avec les technologies d'identification des appareils ;
- Les projets d'évaluation de publicités.

Lorsque des cookies et d'autres objets similaires sont utilisés pour des panels et des enquêtes, les fournisseurs d'échantillons en ligne et les chercheurs doivent se conformer aux lois, réglementations et codes de secteur en vigueur, y compris en ce qui concerne la séparation des activités de recherche et marketing. Dans certaines juridictions, cela implique notamment d'obtenir l'autorisation des membres de panels et des participants aux projets de recherche concernant le stockage de cookies et d'autres objets similaires sur leurs appareils pour la première fois. Les participants aux projets de recherche doivent être informés de ce que sont les cookies et les autres objets similaires et de la fin à laquelle ils sont utilisés. Cette information doit être présentée dans une langue comprise facilement afin que les panélistes et les participants au projet de recherche puissent décider de manière éclairée s'ils souhaitent ou non donner leur permission.

### **3.2.3 Autres méthodes**

Les fournisseurs d'échantillons en ligne sont susceptibles d'utiliser des alternatives à la technologie d'identification d'appareil, aux cookies et aux autres objets similaires s'ils accomplissent des fonctions équivalentes à des niveaux de précision et d'efficacité semblables. Ces méthodes comprennent d'autres solutions technologiques, ainsi que des solutions fondées sur des processus. Toute méthode alternative doit être conforme aux lois et réglementations locales.

Quoi qu'il en soit, la méthode utilisée doit être entièrement documentée, et les résultats du processus de déduplication doivent être fournis au client sur demande.

## **3.3 L'engagement des participants aux projets de recherche doit être calculé et doit faire l'objet d'un rapport**

De nombreux clients sont préoccupés par le fait que les enquêtes en ligne peuvent être particulièrement affectées par des données peu fiables fournies par les participants aux projets de recherche qui ne réfléchissent pas suffisamment avant de répondre aux questions de l'enquête ou qui fournissent délibérément des réponses frauduleuses. Il est important d'identifier ces participants aux projets de recherche afin que leur impact sur les résultats globaux de l'enquête soit minimisé.

La recherche sur la recherche a mis en lumière une vaste gamme de mesures potentielles qui pourraient être utilisées pour identifier les participants inattentifs aux projets de recherche. Voici une liste non exhaustive de ces mesures :

- Temps de réponse à l'enquête ;
- Part des questions laissées sans réponse ;
- Quantité de réponses non substantielles sélectionnées telles que « Je ne sais pas » ou « Refuse de répondre » ;
- Les réponses schématiques aux questions en matrice ou dans un tableau (par ex. réponses linéaires, au hasard, etc.) ;
- La détection des réponses incohérentes, par exemple lorsque des réponses à la fois positives et négatives sont données dans le même lot d'attributs ;
- Les questions « leurres » comprenant plusieurs éléments de réponse, avec des catégories de réponse à faible probabilité ou fictives ;
- Des questions piège telles que « Cochez la case tout à droite » dans une matrice ;
- Réponses appropriées aux questions ouvertes.

Le chercheur chargé de la conception de l'enquête ainsi que la société qui l'héberge se partagent généralement la responsabilité d'identifier les potentiels participants frauduleux ou inattentifs aux enquêtes. Le partage approprié des responsabilités est une question qui doit être négociée entre les deux parties. Il est fortement recommandé d'avoir recours à plusieurs mesures dans la liste ci-dessus et d'utiliser un système de notation qui regroupe ces mesures pour chaque participant, en incluant les scores aux archives de données correspondant au participant au projet de recherche. Les chercheurs et les clients devraient travailler main dans la main afin de déterminer les mesures spécifiques à utiliser, ainsi que le seuil de score pour les enquêtes en deçà duquel les participants aux projets de recherche sont éliminés, le cas échéant. Ces données sont parfois calculées et ces participants aux projets de recherche indésirables sont abandonnés en temps réel durant l'enquête, tandis que dans d'autres cas, cela est fait durant le traitement ultérieur de l'enquête. Le fournisseur d'échantillon en ligne devrait être préparé à remplacer tout participant au projet de recherche dont les données sont considérées comme inacceptables par le client, dans la mesure où les critères de non acceptation ont été précédemment convenus.

Les calculs utilisés ainsi que la méthode de calcul du score global doivent être documentés et partagés avec le client sur demande.

### **3.4 L'identité et les données personnelles des participants aux projets de recherche doivent être protégés.**

Un vaste ensemble de lois, de réglementations et de codes de secteur éthiques exige que toutes les données recueillies auprès des participants aux projets de recherche en ligne soient conservées en toute sécurité et ne soient utilisées qu'à des fins d'étude de marché. Aucune donnée personnellement identifiable ne peut être partagée avec un client sans l'accord du participant au projet de recherche, et lorsque cela est fait, cela doit être effectué en conformité des lois, réglementations et codes de secteur en vigueur. Lorsque le participant accepte de partager ses données avec un client, la responsabilité de conserver les données en toute sécurité et de protéger l'identité des participants au projet de recherche est transférée au client. Le transfert de ces données et le transfert des responsabilités associées doivent être clairement formulés et documentés au moyen d'accords écrits. Lorsque de telles données traversent les frontières nationales, le transfert doit respecter les lois et réglementations telles que la directive européenne de protection des données, un programme agréé tel que la Sphère de sécurité convenue entre les États-Unis et l'UE ou des accords officiels de transfert des données.

### **3.5 Soyez particulièrement vigilants avec les enfants et les jeunes personnes**

Les fournisseurs d'échantillons en ligne doivent veiller à ce qu'aucun enfant ne soit sélectionné pour participer à un projet de recherche sans qu'un parent ou tuteur légal, ou que toute autre personne responsable légalement de l'enfant n'ait donné la permission à cet enfant de participer au projet spécifique pour lequel il ou elle a été sélectionné(e). La définition légale de l'enfant varie substantiellement d'une juridiction à une autre, et le fournisseur d'échantillons doit se conformer aux lois en vigueur dans la juridiction où l'enfant vit. Là où aucune définition nationale spécifique n'existe, les individus âgés de moins de 14 ans doivent être considérés comme des « enfants » et ceux âgés de 14 à 17 ans comme des « jeunes personnes ». Ces tranches d'âges correspondent généralement aux différents stades de développement mental et psychologique. Pour de plus amples renseignements, consultez la directive ESOMAR, [Interviewing Children and Young People](#) (Interviewer les enfants et les jeunes personnes).

Les organisations achetant des échantillons en ligne doivent veiller à garantir que les permissions appropriées ont été obtenues avant d'interroger des enfants.

### **3.6 Avoir recours à de bonnes pratiques de création de questionnaire en ligne**

Malgré près de deux décennies de recherche en ligne et l'existence d'un important organe de recherche sur la recherche à propos de la création des questionnaires en ligne, quelques bonnes pratiques sont couramment admises. Par exemple :

- Plus le questionnaire est long, plus les participants aux projets de recherche sont susceptibles d'abandonner et de potentiellement compromettre la qualité des données. Un certain nombre d'études a révélé une augmentation des comportements tout juste satisfaisants (satisficing) et même d'interruptions après 18 à 20 minutes.
- Les recherches ont également montré qu'une série répétée de questions en matrice en dans un tableau peuvent résulter en des réponses linéaires ou tout autre schéma de réponse.
- Il existe un phénomène connu sous le nom de primauté, lors duquel les questions comprenant un grand nombre de catégories de réponses peuvent pousser les participants aux projets de recherche à choisir des réponses du haut de la liste plus souvent que celles du bas.

En général, les meilleurs questionnaires sont ceux qui comprennent des questions bien conçues, faciles à comprendre et intéressantes, présentées dans un ordre logique avec une liste tout aussi bien conçue de choix de réponse potentiels. Il est dans le meilleur intérêt de toutes les parties — fournisseurs d'échantillons, chercheurs, utilisateurs de données et clients — que les questionnaires respectent les bonnes pratiques du secteur.

### **3.7 Les fournisseurs d'échantillons en ligne doivent être transparents avec les chercheurs et les clients au sujet des sources de leurs échantillons, du processus d'échantillonnage et de ses résultats.**

Si les utilisateurs d'échantillons en ligne sont convaincus que leur échantillon répond à l'objectif, alors les fournisseurs d'échantillons en ligne doivent rendre disponibles les informations concernant le processus de développement des échantillons. Selon l'approche d'échantillonnage spécifique utilisée, cela comprend :

- Une description du cadre d'échantillonnage ou des sources dont l'échantillon est issu (y compris tout sous-traitant utilisé), la façon dont il a été créé ou acquis, et la population cible qu'il est censé représenter.
- La méthode d'échantillonnage utilisée pour sélectionner de potentiels participants aux projets de recherche à partir du cadre d'échantillonnage ou autre, et les moyens employés pour garantir que l'échantillon représente la population cible, y compris les quotas ou les méthodes de panachage d'échantillons utilisées.
- Les critères spécifiques utilisés pour la sélection d'échantillons, tels que les quotas ou les autres critères de filtrage.
- Le(s) type(s) de compensation/rétributions offert(s) aux membres de l'échantillon.
- Lorsque des invitations sont envoyées aux membres de panels, un total du nombre d'unités d'échantillon attirées et sollicitées, le nombre d'e-mails retournés à l'expéditeur (qui peuvent aussi être comptabilisés et supprimés au niveau du panel), le nombre d'entretiens partiels, et le nombre d'entretiens achevés.
- Lorsqu'un routeur ou une méthode d'interception similaire sont utilisés, un total du nombre de participants potentiels au projet de recherche filtrés, les critères spécifiques utilisés, le nombre de participants au projet de recherche correspondant aux critères requis, les règles concernant le nombre d'enquêtes soumises à un participant. Lorsqu'un type de routeur spécifique est connu pour provoquer un certain biais dans la sélection des participants au projet de recherche, cela doit également être documenté.

En outre, les fournisseurs et les utilisateurs d'échantillons en ligne ont des responsabilités de reporting supplémentaires en raison de la variété des méthodes d'échantillonnage utilisées pour les projets de recherche en ligne. Alors que certains panels en ligne sont recrutés en utilisant des méthodes traditionnelles basées sur des probabilités, la plupart ne le sont pas. Les innovations récentes telles que les routeurs en ligne et les avancées en matière de tirage dynamique élargissent la portée sur Internet pour inviter des volontaires à répondre à des enquêtes. En conséquence, la large majorité des échantillons en ligne sont des échantillons pratiques qui n'ont pas les propriétés statistiques que l'on estime nécessaires pour représenter précisément la population cible prévue, entraînant ainsi le risque que les résultats d'une étude contiennent des erreurs conséquentes.

Une pratique commune consiste à imposer des quotas démographiques (principalement en termes d'âge et de sexe) lors de la sélection des échantillons ou des ajustements post-enquête. Un certain nombre d'études a prouvé que ces ajustements étaient parfois insuffisants et que des ajustements supplémentaires basés sur des variables comportementales en lien avec le sujet de l'enquête pouvaient s'avérer nécessaire pour améliorer la précision.<sup>1</sup>

La norme ISO 20252 sur les études de marché, études sociales et d'opinion exige que les chercheurs rapportent à leurs clients « la procédure utilisée pour sélectionner de potentiels participants au projet de recherche à partir du cadre d'échantillonnage ou d'un équivalent, et les moyens employés pour garantir que l'échantillon représente la population cible. » Elle exige en outre que les chercheurs décrivent leurs méthodes de pondération et de projection et fournissent une « évaluation de la mesure dans laquelle l'échantillon représente bien la population cible et les implications associées en termes de qualité de données. » De même, la norme ISO 26362 sur les access panels pour les études de marché, études sociales et d'opinion exige que « le fournisseur d'access panels se mette d'accord avec ses clients sur la conception et les méthodes à utiliser pour tirer des échantillons des access panels pour les enquêtes ou à d'autres fins de recherche. » Elle exige en outre que les méthodes d'échantillonnage utilisées soient rapportées aux clients. ESOMAR et GRBN considèrent ces normes ISO comme de bonnes pratiques que tous les chercheurs devraient respecter.

### **3.8 Les chercheurs utilisant des échantillons en ligne doivent être transparents avec leurs clients.**

Les normes de reporting appropriées pour les projets de recherche basés sur des échantillons en ligne sont semblables à celles des projets de recherche en général. Les données suivantes doivent systématiquement être fournies aux clients :

- Le cadre d'échantillonnage ou un équivalent, les sources et les méthodes d'échantillonnage utilisées
- Les dates du travail de terrain
- La durée d'enquête moyenne ou médiane
- Le nombre total d'entretiens entièrement complétés
- Tout quota utilisé ou toute autre spécification utilisée dans le cadre de la sélection de l'échantillon

---

<sup>1</sup> Pour en savoir plus, consultez le rapport du groupe de travail sur l'échantillonnage hors probabilité, disponible sur le site de l'American Association for Public Opinion Research ([www.aapor.org](http://www.aapor.org)). Un récapitulatif du rapport ainsi que des critiques de plusieurs experts dans le domaine a été publié dans *Le Journal des statistiques et de la méthodologie d'enquête*, Volume 1, Numéro 2, novembre 2013.

- Le questionnaire et les autres documents de collecte des données pertinents
- Un total du nombre de participants à l'enquête dont l'identité a été validée
- Une description des méthodes de déduplication utilisées et le nombre de réponses supprimées dans ce cadre
- Les calculs de l'engagement des participants aux projets de recherche utilisés et un total de tous les participants au projet de recherche supprimés ou remplacés en raison d'un mauvais comportement durant l'enquête
- Les informations d'exclusion
- Les taux de participation<sup>2</sup> (dans la mesure du possible) et les méthodes utilisées pour les calculer
- Taux d'achèvement
- Si une partie du projet ou le projet complet a été sous traité et, le cas échéant, auprès de quelles organisations.

### **3.9 Collecte passive de données**

Les méthodes de collecte passive de données impliquent souvent des données considérées comme des données personnelles. Les sources de données passives comprennent des données de navigation web, des données relatives aux applications, à la fidélisation, à la géolocalisation, aux réseaux sociaux, et générées par/obtenues par des appareils mobiles. Une grande partie de ces données peut être combinée à des données d'enquête.

Toutes les méthodes de collecte des données passives utilisées doivent être conformes aux lois et réglementations en vigueur. Tout comme avec les données personnelles, les chercheurs basés dans de nombreuses juridictions devront établir une base légale claire pour utiliser et traiter ces données, y compris en ce qui concerne leur utilisation pour les activités d'échantillonnage et pour obtenir le consentement des individus concernés.

### **3.10 Se conformer à l'ensemble des lois, réglementations et codes de conduite du secteur pertinents**

Il est essentiel qu'à la fois les fournisseurs et les acheteurs d'échantillons en ligne soient conscients et adhèrent strictement à toutes les réglementations régionales, nationales et locales pertinentes, ainsi qu'aux codes de conduite du secteur ou aux habitudes culturelles qui sont susceptibles d'établir des normes plus hautes que les exigences légales.

## **4 CONSIDÉRATIONS SPÉCIALES**

À mesure que les méthodes et processus de recherche continuent de changer et d'évoluer, de nouveaux défis en termes de qualité des échantillons émergent. Malgré ces changements, les responsabilités des chercheurs utilisant des méthodes d'échantillonnage en ligne restent identiques.

### **4.1 L'exclusion par catégorie et autres types d'exclusion**

Pour certaines études, les chercheurs peuvent désirer exclure certains participants aux projets de recherche s'ils ont récemment participé à une enquête afin d'éviter les partis pris. Cela peut notamment être le cas pour les participants aux enquêtes ayant récemment

---

<sup>2</sup> La présente directive reconnaît que l'utilisation croissante de l'échantillonnage aléatoire et des routeurs rend le calcul des taux de participation difficile voire impossible. Jusqu'à ce qu'une bonne pratique généralement acceptée émerge, les chercheurs devraient mentionner cette difficulté dans leurs rapports à leurs clients.

achevé une enquête à propos d'un produit semblable, utilisant une méthode semblable (par ex. étude de prix ou test de concept) ou tout simplement ayant répondu à un très grand nombre d'enquêtes. Étant donné la variété des sources d'échantillons désormais utilisées, il est difficile voire impossible pour un fournisseur d'échantillons de connaître l'historique complet de chaque participant à un projet de recherche dans un échantillon donné. Lorsqu'ils répondent à une requête d'exclusion, les fournisseurs d'échantillons doivent révéler la ou les méthodes spécifiques utilisées et fournir des preuves de leur efficacité.

## **4.2 Mobile**

La prolifération des appareils mobiles (téléphones classiques, smartphones et tablettes) dans le monde entier a créé de nouvelles opportunités d'entrer en contact avec des participants potentiels aux projets de recherche. En ce qui concerne l'échantillonnage, deux problèmes ont émergé.

### **4.2.1 Les participants fortuits aux projets de recherche sur mobile**

Il a été prouvé qu'un nombre croissant de participants aux projets de recherche choisissent de répondre aux enquêtes en ligne en utilisant un appareil mobile. Les recherches sur la recherche actuelles suggèrent qu'en moyenne 20 à 30 % des participants aux projets de recherche sont susceptibles de répondre en utilisant un smartphone ou un autre appareil mobile. On peut raisonnablement penser que ce phénomène va prendre de l'ampleur avec le temps. En outre, la recherche sur la recherche a également démontré que ces participants aux projets de recherche étaient souvent démographiquement différents des autres participants aux projets de recherche : plus jeunes, d'origines plus diverses, plus susceptibles d'être des hommes que des femmes, etc. En parallèle, les enquêtes en ligne ne sont pas toujours optimisées pour les plus petits écrans de beaucoup d'appareils mobiles, ce qui entraîne parfois des taux d'achèvement plus bas ou des réponses aux enquêtes biaisées.

Il existe généralement trois catégories de réponse à cet usage croissant des appareils mobiles par les participants : (1) ne pas autoriser la participation sur un appareil mobile ; (2) autoriser les participants à répondre, mais ne pas adapter l'enquête aux limites des appareils mobiles ; et (3) optimiser la présentation de l'enquête pour les mobiles, bien que les bonnes pratiques dans ce domaine restent indéterminées.

Chacune de ces trois approches est susceptible d'entraîner une certaine forme de déformation. Ainsi, les chercheurs devraient inclure les informations suivantes à leurs rapports aux clients :

- Si les participants au projet de recherche étaient autorisés à répondre sur un appareil mobile
- Le nombre d'enquêtes terminées, entièrement ou en partie, sur un appareil mobile
- Les implications en termes de représentativité de la population cible par l'échantillon

### **4.2.2 Les nouvelles sources d'échantillon pour les études de marché mobiles**

La deuxième sorte d'utilisation des appareils mobiles dans le cadre de la recherche correspond à la catégorie que l'on pourrait appeler la « recherche sur mobile pure ». Cette catégorie se base sur les capacités et les schémas d'utilisation uniques de ces appareils et comprend la recherche instantanée, l'utilisation des informations de géolocalisation, et de nouvelles formes de collecte de données passive. Ces méthodes nécessitent habituellement le recours à des échantillons d'utilisateurs de mobiles connus. De nombreux fournisseurs d'échantillons en ligne ont créé des panels mobiles. Les fournisseurs d'échantillons en ligne utilisant l'échantillonnage aléatoire et d'autres formes de tirage dynamique ont commencé à filtrer les utilisateurs de mobile et ont ajouté les sources de participants sur mobile uniquement, telles que les réseaux de publicité sur mobile. Dans tous les cas, les fournisseurs d'échantillons comme les chercheurs doivent respecter le niveau de précaution et la qualité du reporting détaillés dans la section 3.8 de la présente directive.

### 4.3 Automatisation des enquêtes et de l'échantillonnage

La demande croissante concernant l'automatisation des services de recherche soulève de nouvelles difficultés en termes de gestion et de garantie de la qualité des échantillons. Les caractéristiques clés de ces offres comprennent l'automatisation des processus d'échantillonnage (y compris les échanges de sources d'échantillons en ligne) et un délai de livraison rapide. Les dispositifs de garantie de la qualité des échantillons décrits dans la présente directive ne sont pas toujours intégrés à des services de recherche maison et un manque de transparence peut s'avérer. Les chercheurs envisageant l'utilisation de tels services devraient exiger la transparence et étudier leurs options soigneusement, comme ils le feraient en sélectionnant un fournisseur d'échantillons en ligne classique. Les concepts présentés dans cette directive et d'autres sources, y compris [28 Questions to Help Buyers of Online Samples](#) (28 questions pour aider les acheteurs d'échantillons en ligne), peuvent être utilisés comme bonne base de départ pour réaliser une telle évaluation.

## 5 L'ÉQUIPE DU PROJET

Reg Baker, Coprésident, Consultant auprès du Professional Standards Committee d'ESOMAR, Marketing Research Institute International

Peter Milla, Coprésident, Consultant technique chez CASRO, Peter Milla Consulting

Pete Cape, Survey Sampling International

Mike Cooke, GfK

Melanie Courtright, Research Now

George Terhanian, The NPD Group

## 6 RÉFÉRENCES

[28 Questions to Help Buyers of Online Samples \(28 questions pour aider les acheteurs d'échantillons en ligne\)](#)

[Directive sur la conduite des études de marché sur mobile, ESOMAR](#)

[Directives sur les études de marché mobiles, GRBN](#)

*The Impact of Digital Fingerprinting and Data Verification on Data Quality (L'impact de l'empreinte numérique de la vérification des données sur la qualité des données)*, Les Fondamentaux de la qualité 2.0 de l'Advertising Research Foundation, par Courtright et Pashupati, 2014

[Interviewing Children and Young People \(Interviewer les enfants et les jeunes personnes\)](#)

ISO 20252 – Études de marché, études sociales et d'opinion

ISO 26362 – « Access panels » pour les études de marché, études sociales et d'opinion