



Nouveaux enjeux de la donnée et méthodes d'échantillonnage

*Journée d'études sur l'échantillonnage,
PUD-AMU & Progedo, 8 novembre 2022*

Frédéric Gonthier

Sciences Po Grenoble, Université de Grenoble Alpes, Pacte-CNRS

frederic.gonthier@iepg.fr



Deux objectifs

1. Retour réflexif sur les méthodes d'échantillonnage/modes de collecte et les nouveaux enjeux de la donnée en SHS
 - L'« erreur totale » (paradigme *Total Survey Error*)
 - Son application au recueil des données en SHS
2. Prise de recul sur l'accompagnement à la production et analyse de données innovantes au sein des PUD
 - Ce que la Big Data fait aux SHS
 - Les biais liés au recrutement/recueil par internet

1. Méthodes d'échantillonnage et modes en collecte en SHS

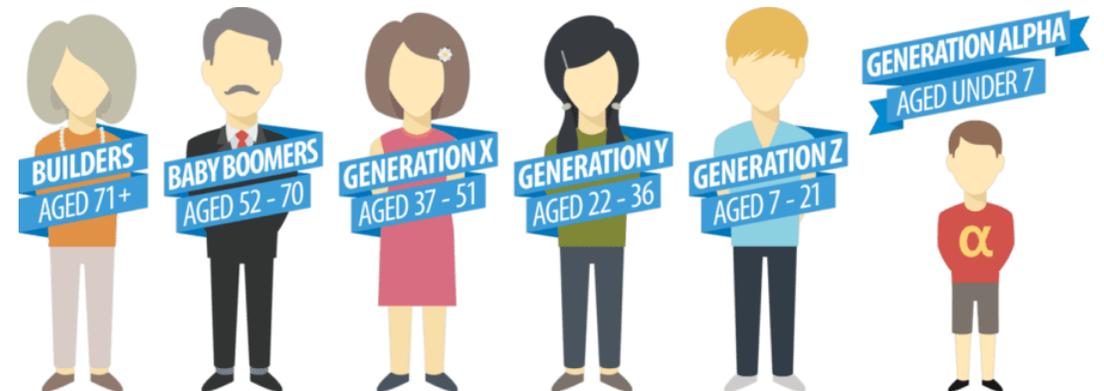
Pourquoi et comment échantillonner ?

Préambule : Les données ne sont pas données

- Les données d'enquête sont « construites » : elles résultent d'un travail d'élaboration, dont les méthodes d'échantillonnage font partie intégrante
 -  « Construction » ne signifie pas « invention », mais fruit des dispositifs d'enquêtes pour enregistrer le « réel » (procédures d'enregistrement, de codification et de traitement des informations)
- Les données sont « coûteuses » : la réalisation d'une enquête engage un ensemble d'opérations onéreuses (reproduction et diffusion des questionnaires, rémunération des enquêteurs, saisie des réponses...)
- Les données sont « conquises » : la conception et l'analyse d'une enquête quantitative visent à rompre avec la connaissance spontanée provenant de nos expériences ordinaires
- Pour aller plus loin: Alain Desrosières, [« La statistique, outil de libération ou outil de pouvoir ? »](#), *Prouver et gouverner*, La Découverte, 2014, 70-84

La construction des catégories statistiques

- Plusieurs exemples en statistique publique...
 - Les chiffres du chômage (au sens du BIT vs halo du chômage)
 - Le taux de pauvreté (pauvreté absolue vs relative, combien de % du revenu médian ?)
 - L'indice des prix à la consommation (quels biens ? comment intégrer les évolutions ?)
- ... mais aussi dans les enquêtes par sondage ...
 - Au moment de la définition de la population d'enquête
 - Au moment de l'analyse des données (regroupement de catégories)



Les données sont (re)traduites

ÉTAPE	ACTEUR
1. Questionnaire	STATISTICIEN
2. Champ de l'enquête, échantillon et plan de codage	
3. choix des enquêtés	Enquêteur
4. questionnements	
5. Interprétation des questions	ENQUETE
6. Réponses	
7. Inscription des réponses	Enquêteur
8. Nomenclatures	STATISTICIEN
9. Chiffrement (codage)	CODEUR
10. Saisie informatique	OPERATEUR DE SAISIE
11. Nettoyage et redressement	Informaticien
12. Demande de tri	STATISTICIEN
13. Tableaux statistiques	Informaticien
14. Analyse de l'information	STATISTICIEN
15. Diffusion (restreinte)	SERVICE DES PUBLICATIONS
16. Retraduction	« MASS MEDIA » ET JOURNALISTES
17. Diffusion (large)	
18. Réception	« PUBLIC »
19. Utilisation	

Source : Dominique Merllié, « La construction statistique », in Champagne P. et al., *Initiation à la pratique sociologique*, Paris, Dunod, 1989, p. 157

Méthodes d'échantillonnage et représentativité (1/2)

- Par convention, les données dites de population (N , aussi appelé univers, population mère...) sont des données exhaustives (i.e., elles proviennent de tous les cas compris dans l'univers)
- L'échantillon (n) est un sous-ensemble tiré de la population mère. Il vise à former un « modèle réduit » de la population, intégrant son hétérogénéité/variabilité selon des critères définis a priori
- N et n étant « indépendants », lors de l'échantillonnage



Vigilance sur la taille de l'échantillon : gains en précision + faibles quand $n > 1000$



Vigilance sur le taux de sondage : plus n/N est petit (i.e., $1/20$), plus l'exhaustif devient avantageux



Méthodes d'échantillonnage et représentativité (2/2)

- La notion de représentativité statistique est elle-même un construit social, qui a fait l'objet de vives controverses au XIXème et début XXème
 - Partisans de l'approche aléatoire vs partisans de l'approche par choix raisonné
 - Légitimation progressive de la méthode représentative (travaux pionniers d'Anders Kiaer, 1895 et Arthur Bowley, 1906 et 1912 ; commission de l'Institut international de statistique, 1924-1925)
 - Bowley (1926) et Jerzy Neyman (1933) jettent les bases de l'échantillonnage aléatoire stratifié
- Représentativité n'est pas égal à miniaturisation
 - Un échantillon n'est pas forcément un parfait décalque (cela peut être fortuit), mais doit surtout respecter les règles de construction aléatoire
 - Porter l'attention sur la qualité de la base de sondage autant que sur le calage brut de la population

A quelle étape d'une enquête quantitative se poser la question de l'échantillonnage ?

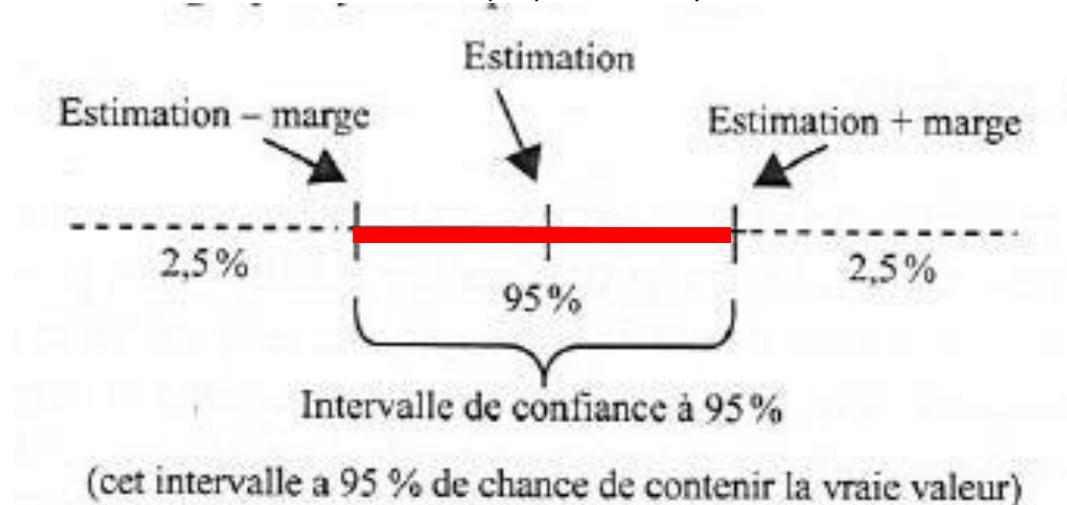
1. Etude de la commande, élaboration du cahier des charges, définition des questions de recherche et de la population d'enquête
2. Construction du plan de sondage:
 - Définition de la base de sondage (listes sur lesquelles l'échantillon est constitué)
 - **Choix de la méthode d'échantillonnage** (mode de sélection des individus) et de la taille de l'échantillon (taux de sondage)
 - Choix du mode de collecte
3. Élaboration du questionnaire (pré-test)
4. Collecte des données (passation, nettoyage, codage, documentation)
5. Exploitation et analyse statistique
6. Production des rendus (rapports, synthèses, infographie)
7. Diffusion et valorisation

Pourquoi échantillonner ?

- Les méthodes d'échantillonnage visent à compenser les limites de l'approche exhaustive
 - Les données sur une population d'intérêt sont impossibles à obtenir en intégralité
 - Même si c'est le cas, se concentrer sur un échantillon permet d'être plus vigilant sur le dispositif, passer plus de temps, obtenir des informations de meilleure qualité...
- L'inférence désigne la production d'information sur une population à partir de l'observation d'un échantillon, grâce à des règles « d'extrapolation »
 - L'inférence n'est pas tant une mesure du « vrai chiffre » que de l'incertitude qui accompagne un chiffre probable
- La théorie des sondages permet d'établir les limites de l'incertitude, sous la forme de conclusions probables
 - Les résultats ont une portée générale, mais avec une marge d'erreur.

Marges d'erreur et intervalle de confiance (1/2)

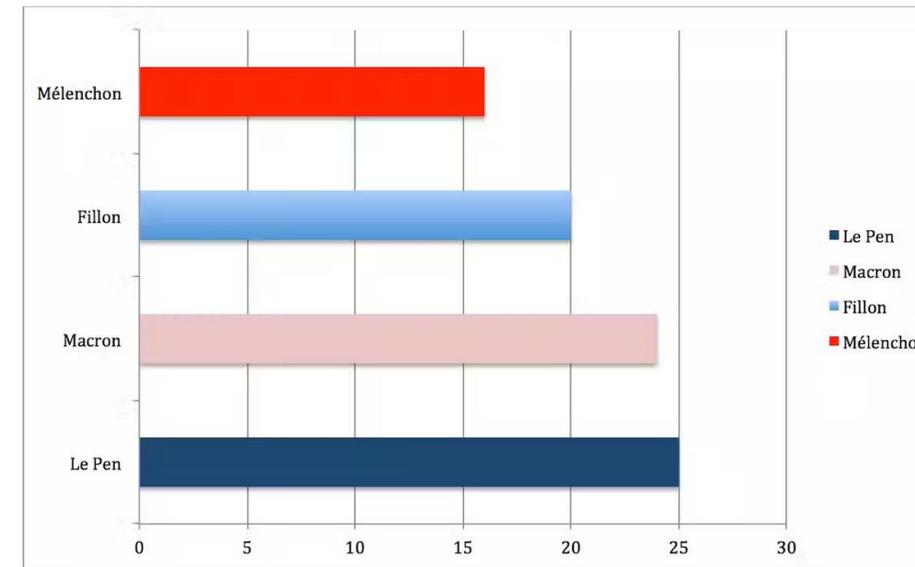
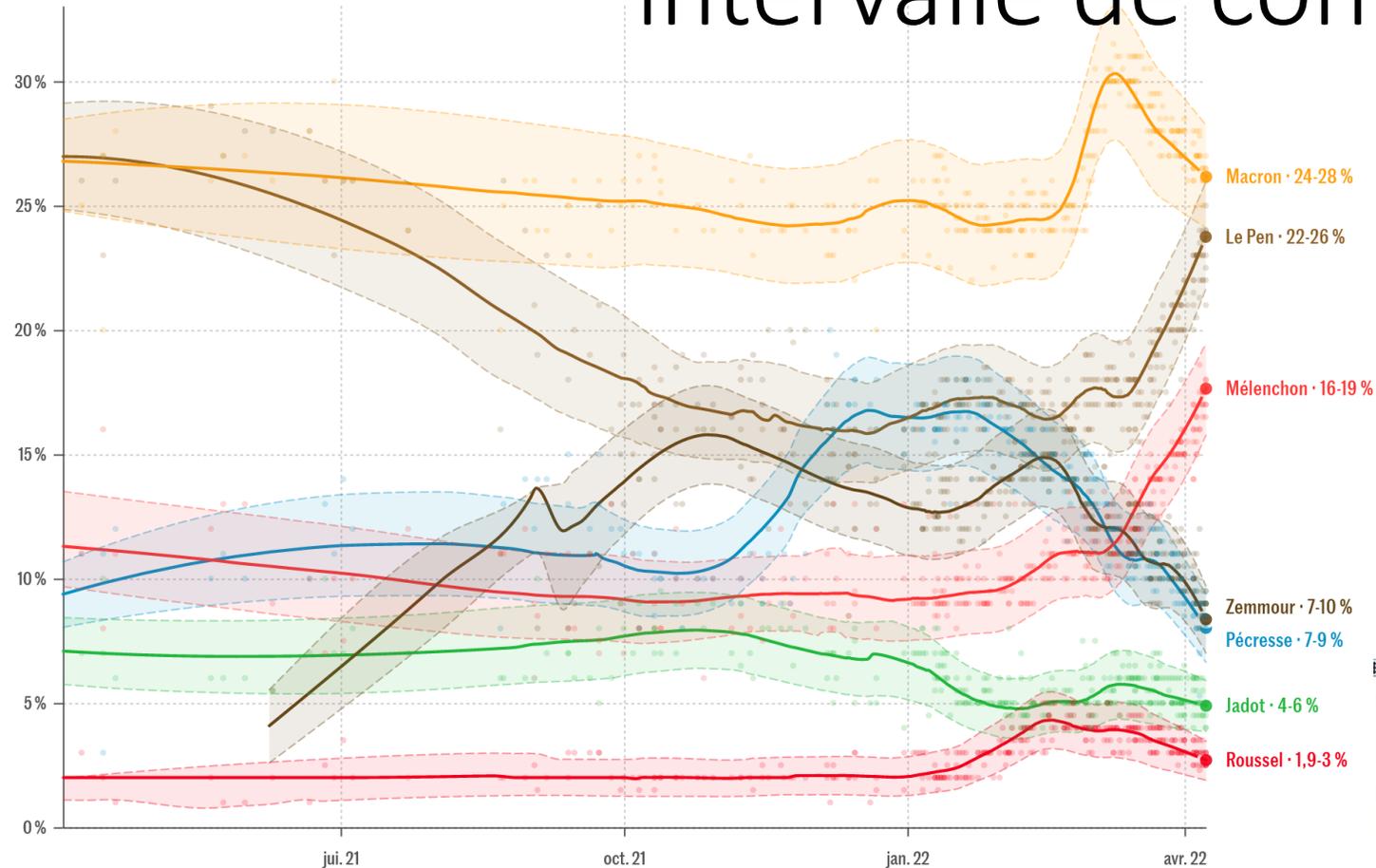
- Marge d'erreur : mesure statistique qui rend compte de la différence entre l'estimation trouvée dans l'échantillon, et la "vraie" valeur qu'on obtiendrait si on pouvait recenser toute la population.
- Sert à évaluer si l'on a obtenu notre estimation par hasard ou non, et si l'on trouverait la même estimation si on réalisait plusieurs tirages statistiques sur des échantillons de même taille et de même structure.
- L'intervalle de confiance (IC) est simplement le double de la marge d'erreur :



L'usage général en sciences sociales est qu'il y ait au moins 95% de chances que la relation ne soit pas fortuite

= 95% sûr que notre estimation trouvée dans l'échantillon se situe dans l'intervalle de confiance si l'on veut extrapoler à la population de référence
= 5% de chances que la vraie valeur ne soit pas dans cette intervalle

Marges d'erreur et intervalle de confiance (2/2)



Presitrack Opinionway 7 avril (2^e « marges d'erreur » les plus FORTES pour les 4 can

	14	14,5	15	15,5	16	16,5	17	17,5	18	18,5	19	19,5	20	20,5	21	21,5	22	22,5	23	23,5	24	
Le Pen																						
Macron																						
Fillon																						
Mélenchon																						

https://www.lemonde.fr/les-decodeurs/article/2022/02/08/election-presidentielle-2022-que-montrent-les-principaux-sondages_6112830_4355770.html

<https://theconversation.com/sondages-dintention-de-vote-donnez-nous-les-marges-d-erreur-75992>

La marge d'erreur pour une proportion donnée

Trois éléments à prendre en compte:

- La valeur de l'estimation concernée ici la proportion notée pr
- La taille de l'échantillon notée n
- Le seuil d'erreur accepté (appelé aussi le niveau de confiance). Par convention on se base sur un seuil d'erreur à 5% (soit un niveau de confiance de 95%)

$$me_{pr} = 1.96 \sqrt{\frac{pr(1 - pr)}{n}}$$

me = marge d'erreur, pr =proportion, n =taille de l'échantillon



Plus la taille de l'échantillon est grande, plus le n au dénominateur est élevé, donc plus on obtient mathématiquement une petite marge d'erreur

Plus la proportion est proche de 0.5 (50%), plus le numérateur est élevé, donc plus on obtient mathématiquement une grande marge d'erreur

Méthodes d'échantillonnage

Deux familles de méthode d'échantillonnage (deux grandes façons de recueillir un échantillon aléatoire)

1. Méthodes probabilistes :

- on dispose d'une liste (exhaustive) des individus apparentés à la population mère,
- leur probabilité d'être sélectionnée est connue (i.e., peut être calculée)

2. Méthodes non probabilistes/empiristes/empiriques :

- on ne dispose pas d'une liste complète des individus apparentés à la population mère,
- leur probabilité d'être sélectionnée n'est pas connue (i.e., ne peut pas être calculée)

Les méthodes probabilistes (1/4)

1. *Echantillonnage aléatoire simple* (avec tirage aléatoire, principe de la loterie) ou *systematique* (avec un pas de tirage, par exemple sur une liste d'annuaire téléphonique)
 - Idéal mais suppose une base de sondage complète et accessible pour la recherche
 - + de chances d'avoir une base de sondage exhaustive quand il s'agit d'une sous-population localisée (par exemple, fichiers clients/adhérents/salariés...), mais une enquête exhaustive peut alors être + avantageuse qu'un échantillon

Les méthodes probabilistes (2/4)

2. *Echantillonnage stratifié avec effectifs proportionnels à ceux de la population (= un échantillon aléatoire simple dans chaque strate) :*

- Par exemple, on divise la population de la France métropolitaine par région, et on sélectionne un effectif proportionnel d'individus ayant cette région comme résidence principale
- Pertinent quand la population n'est pas distribuée de façon égale sur une ou plusieurs strates d'intérêt
- Permet de s'assurer que chaque strate est bien représentée
- Plus les strates sont discriminantes, plus il y a réduction de la marge d'erreur

Les méthodes probabilistes (3/4)

3. *Echantillonnage stratifié avec suréchantillonnage de certaines strates*
- Par exemple, pour une enquête sur la détention d'actifs financiers, il est cohérent d'améliorer la représentativité des catégories de revenus supérieures où se concentrent les actifs financiers
 - Pour une enquête sur la fréquentation des cinémas, il est pertinent de segmenter par type d'habitat (urbain/péri-urbain/rural), voire également de surreprésenter l'habitat urbain (où + d'opportunités d'accès au cinéma)
 - Permet d'optimiser les marges d'erreur et de surreprésenter une strate particulièrement intéressante
 - on surreprésente pour être plus précis en allant chercher la variabilité statistique au niveau où elle se trouve
 - Nécessite de rétablir ensuite les strates à leur poids réel (par redressement)

Les méthodes probabilistes (4/4)

4. *Echantillonnage par grappes*

- Peut être à *un degré* quand une sous-population toute entière est présente dans chaque grappe. Par exemple, tous les individus de ménages sélectionnés aléatoirement sont interrogés, ou tous les individus vivant dans une zone choisie aléatoirement (sondage aréolaire)...
- Ou à *deux degrés* quand seul un échantillon figure dans chaque grappe
- Permet de mieux représenter l'effet de grappe, typiquement le regroupement « naturel » des individus en grappes de ménages (au sens de personnes occupant le même logement) ou en grappes géographiques (communes, IRIS...)
- Permet de faire une sélection aléatoire à partir d'une base de sondage agrégée (sans avoir à disposer d'une base de sondage individuelle)
- La marge d'erreur étant plus forte qu'avec le sondage aléatoire simple (effet de grappe), plutôt utiliser de petites grappes

Les méthodes non probabilistes

1. Echantillonnage de *convenance*

- Peu coûteux mais souvent de nombreux biais, notamment de couverture

2. Echantillonnage par *boule de neige*

- Plus coûteux que l'échantillon de convenance
- Des probabilités d'inclusion des répondants difficiles à contrôler ex ante
- Et à corriger ex post (par exemple, enquête VICO)

3. Echantillonnage par *quotas* (simples ou croisés – par exemple âge x diplôme)

- Permet de reproduire les caractéristiques de la population mère selon certains critères d'intérêt



La représentativité sur certains critères ne garantit en rien la représentativité sur d'autres

La recherche des quotas ne doit pas détourner des enjeux liés à la qualité des bases de sondage et des modes de recueil

- Pour aller plus loin, M. Lejeune, *La singulière fabrique des sondages d'opinion*, Paris, L'Harmattan, 2021

2. Des méthodes d'échantillonnage aux modes en collecte en SHS

Vers une approche Total Survey Error

Les grands modes de collecte

- Deux critères : le mode d'administration / le mode de recueil

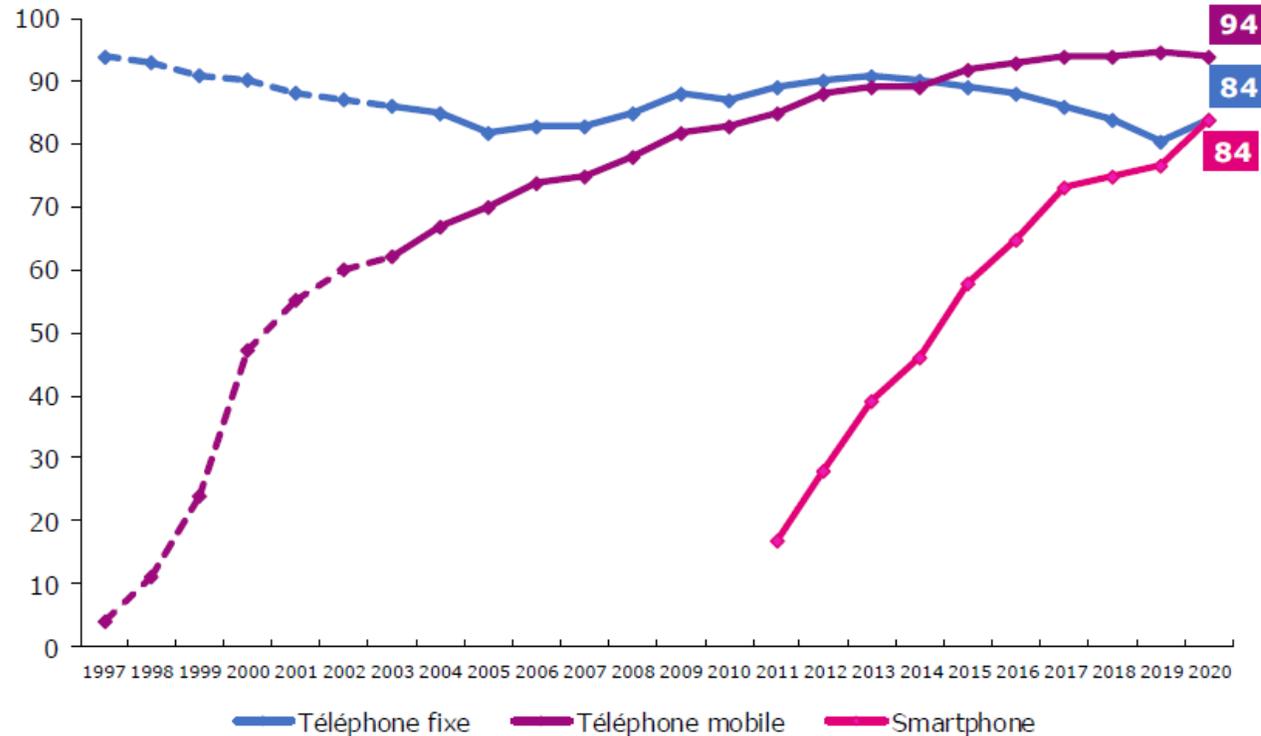
	Paper & Pencil	Computer Assisted
Enquêteur en face-à-face	PAPI	CAPI
Enquêteur à distance	PAPI	CATI
Pas d'enquêteur (auto-administré)	QAA par voie postale	CASI (CAWI / CAMI)

Avantages et inconvénients en termes méthodologiques

	PAPI	CAPI	CATI	QAA P&P	QAA CAWI/CAMI
Coût	Elevé	Elevé	Moyen	Faible	Faible
Délais	Longs	Moyens	Courts	Longs	Court- Moyens
Taux de réponse	Élevé	Élevé	Moyen	Faible	Faible
Longueur du questionnaire	Oui	Oui	Non	Non	Moyen
Richesse/varieté du questionnaire	Non	Oui	Non	Non	Oui

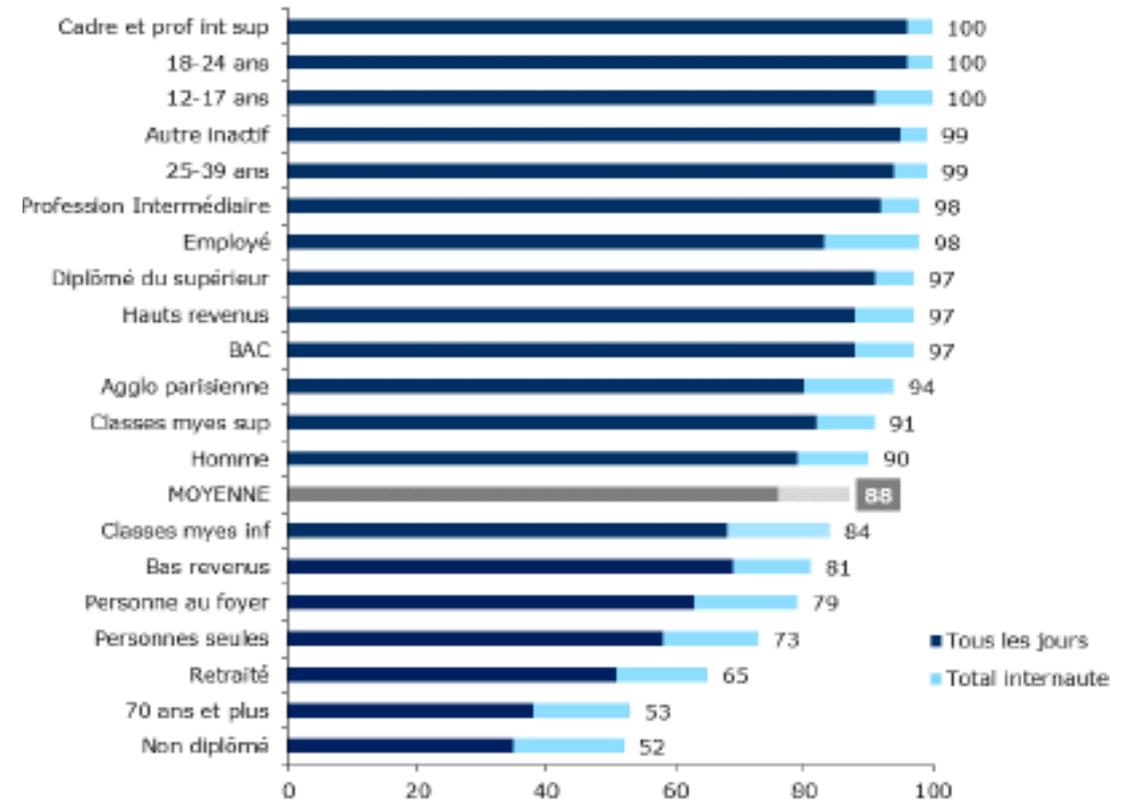


CAWI-CAMI : des disparités dans le taux d'équipement et la fréquence d'utilisation des appareils



Source : Crédoc, [Baromètre du numérique 2021](#)

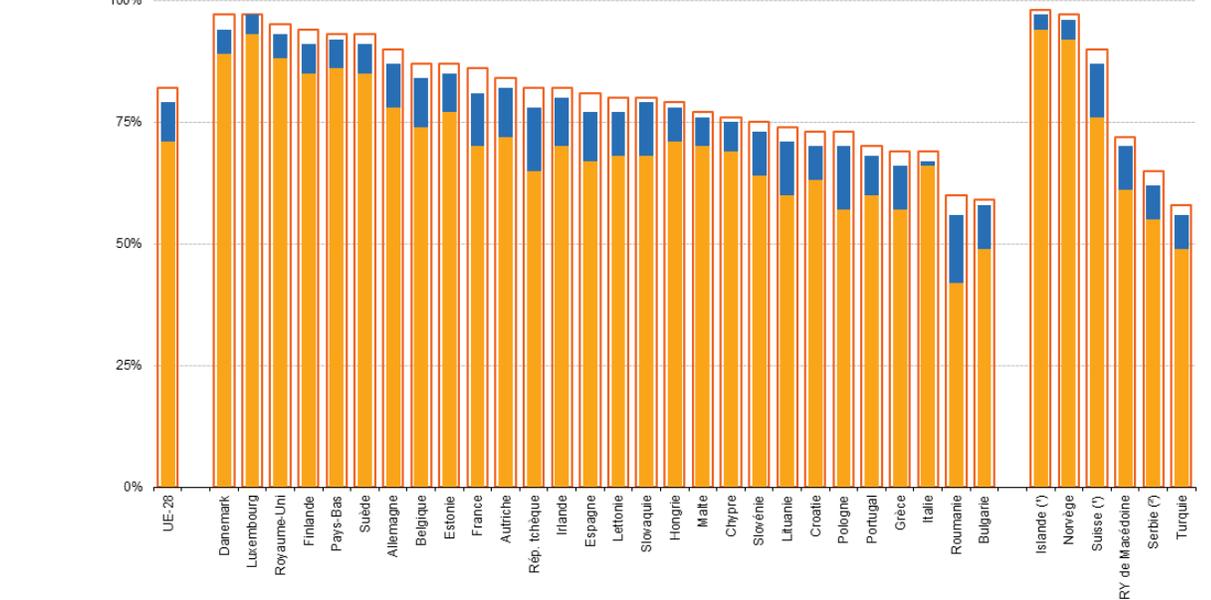
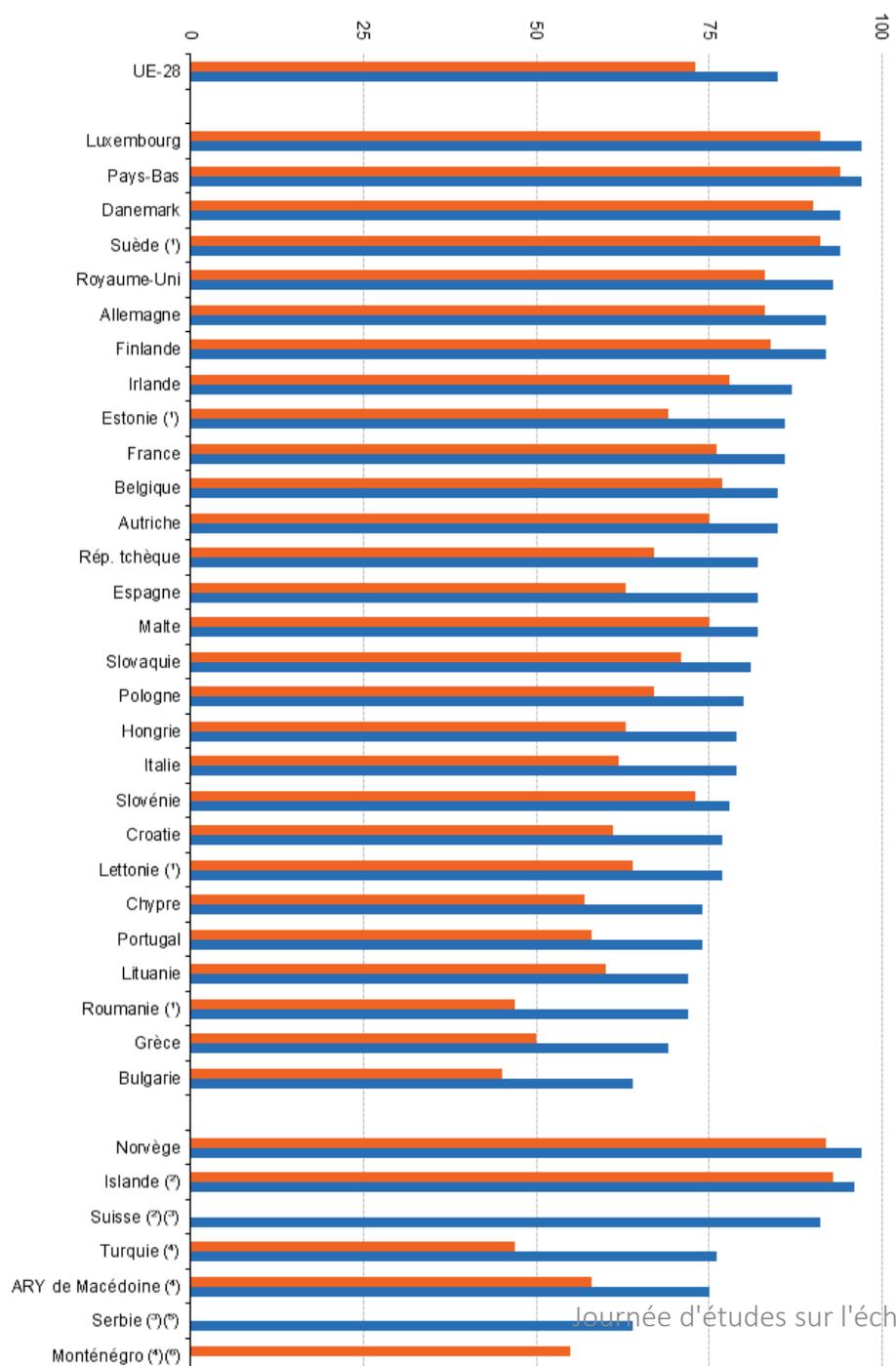
Graphique 30 – Groupes les plus et les moins internautes :
- Champ : ensemble de la population de 12 ans et plus, en % -



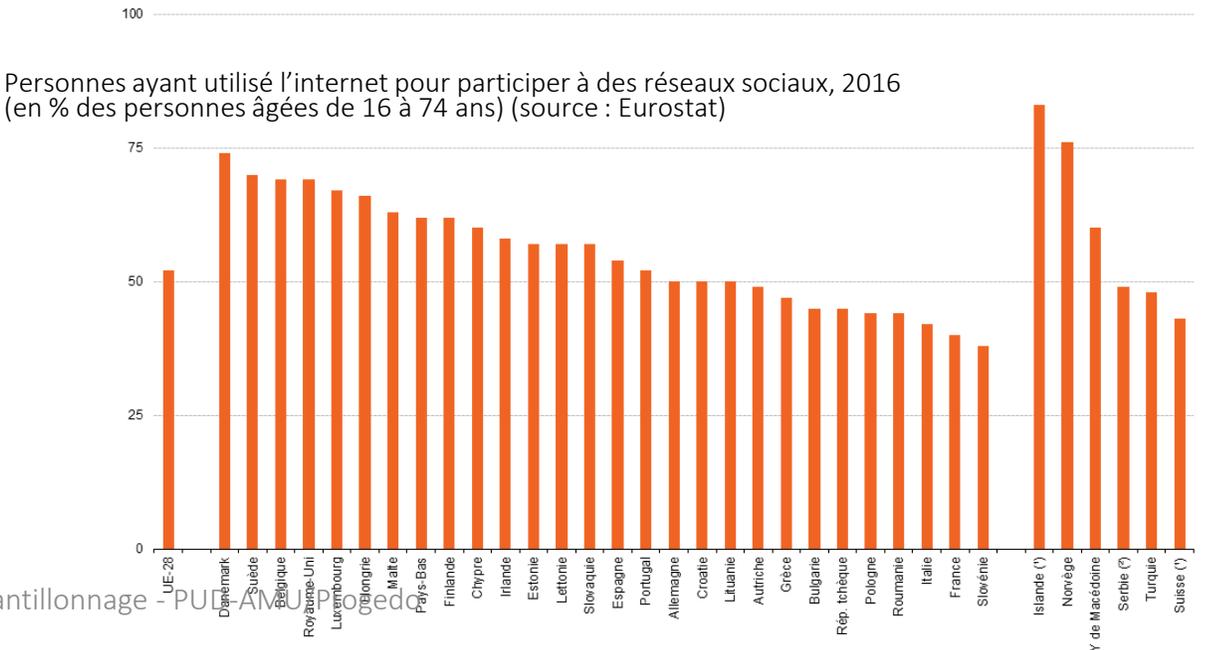
Source : Crédoc, *Conditions de vie et aspirations 2017*

Accès des ménages à l'internet, 2011 et 2016 (en % de l'ensemble des ménages)

■ 2011 ■ 2016
 (*) Rupture de série.
 (†) 2014 au lieu de 2016.
 (‡) 2011 : non disponible.
 (§) 2012 au lieu de 2011.
 (¶) 2015 au lieu de 2016.
 (¶) 2016 : non disponible.
 Source : Eurostat (code des données en ligne : isoc_cl_in_h)



Fréquence d'utilisation d'internet, 2016 (en % des personnes âgées de 16 à 74 ans) (source : Eurostat)



Personnes ayant utilisé l'internet pour participer à des réseaux sociaux, 2016 (en % des personnes âgées de 16 à 74 ans) (source : Eurostat)

Quatre grands types d'erreur statistique

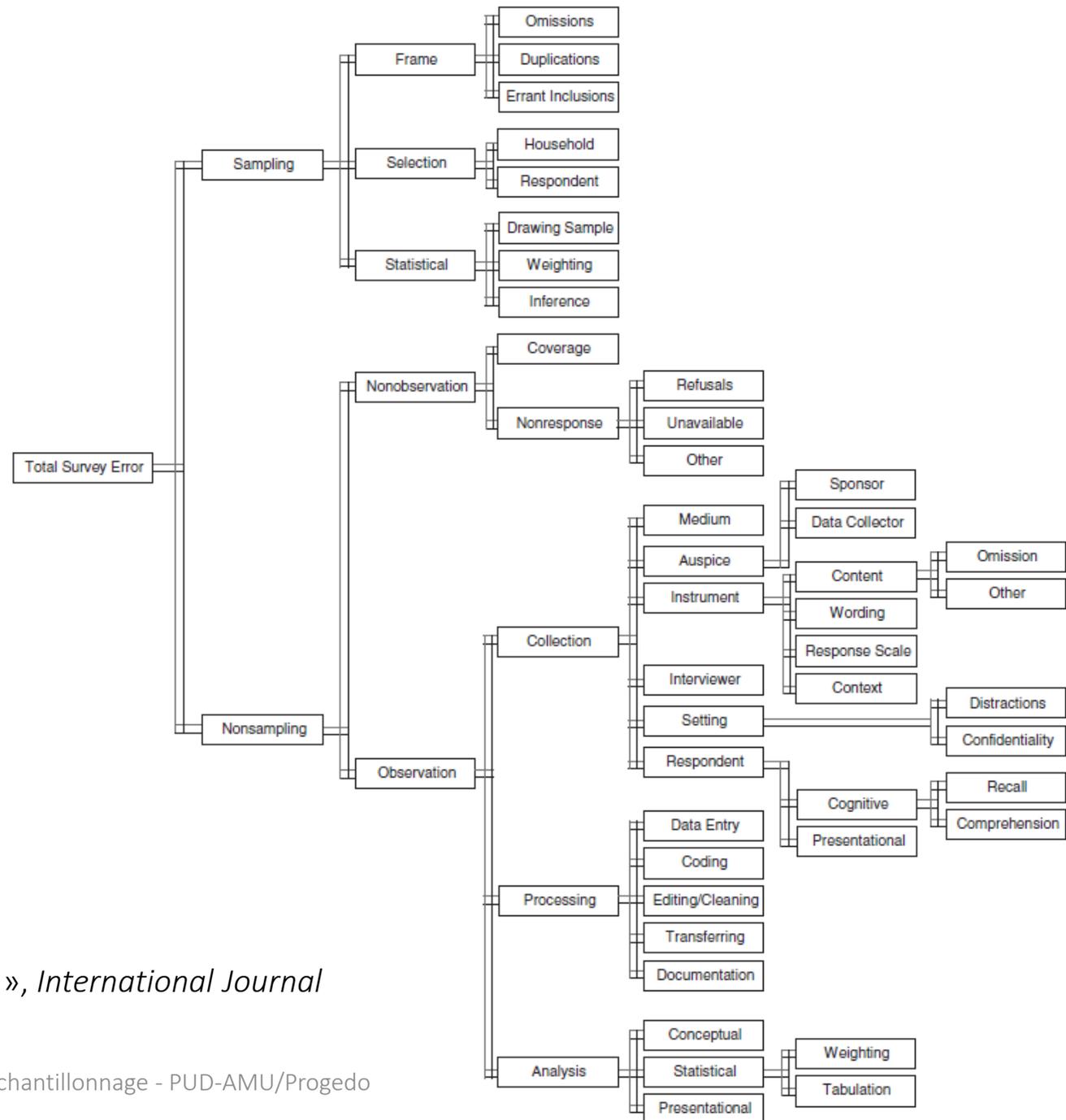
1. *Biais de couverture* : la base de sondage ne correspond pas forcément à la population cible. Risque de sous-dénombrement (unités avec une probabilité nulle d'être sélectionnées, par exemple utilisateurs occasionnels d'internet)
2. *Biais d'échantillonnage* : les estimations se fondent des individus sélectionnés pour faire partie d'un échantillon, et non sur la population dans son ensemble (important si répondants recrutés online)
3. *Biais de non-réponse* : certains individus ne répondent pas (taux de réponse problématique pour les *access panels*)
4. *Biais de mesure* : erreurs liées aux effets de la situation d'enquête et au mode de collecte (désirabilité sociale, *satisficing* indiquant faible implication, *agreement* indiquant faible effort cognitif...), mais aussi au questionnaire (design, formulation des questions, ordre, type d'échelle de mesures...)

L'« erreur totale »

Chaque cartouche correspond à une source/composante d'erreur possible, pour une étape donnée du dispositif d'enquête par sondage

Les deux lignes associées à chaque cartouche renvoient aux deux sources d'erreur statistique possibles :

- la variance (ou l'erreur standard) qui est aléatoire et n'a pas d'effet attendu sur les valeurs estimées ;
- le biais proprement dit (ou l'erreur systématique) qui a un effet directionnel et peut altérer les valeurs estimées



Source : T. Smith, « [Refining the Total Survey Error Perspective](#) », *International Journal of Public Opinion Research*, 23-4, 2011, 464–484

De l'erreur totale à l'enquête multimode

- Collecte multimode : différents modes de recueil (face-à-face, téléphone, papier, internet) intégrés dans la phase de collecte
- Objectifs :
 - Répondre à la baisse tendancielle des taux de réponse
 - Diminuer les coûts induits par la collecte
 - Prendre au sérieux les forces/faiblesses inhérentes à chaque mode de recueil
- En théorie : compenser les faiblesses de chacun des mode de recueil en les combinant
- En pratique :
 - Réduire le biais de couverture en combinant CAPI (un mode performant mais coûteux) avec CAWI (un mode à faible coût marginal mais ne couvrant pas toute la population d'intérêt)
 - Réduire le biais de non-réponse en offrant CAPI/CATI et CAWI/CAMI (un mode auto-administré alternatif pour les non-répondants)
 - Réduire le biais de mesure en proposant des mode spécifiques (QAA) pour un sous-ensemble de questions sensibles...
- Pour aller plus loin :
 - T. Razafindranovona, « [La collecte multimode et le paradigme de l'erreur d'enquête totale](#) », Séminaire de Méthodologie Statistique du département des méthodes statistiques INSEE, 22 janvier 2013
 - G. Brillhault et L. Castell, « [Le multimode : enquêter par différents moyens](#) », Tuto@Mate, webinaire 22 novembre 2021 Journée d'études sur l'échantillonnage - PUD-AMU/Progedo

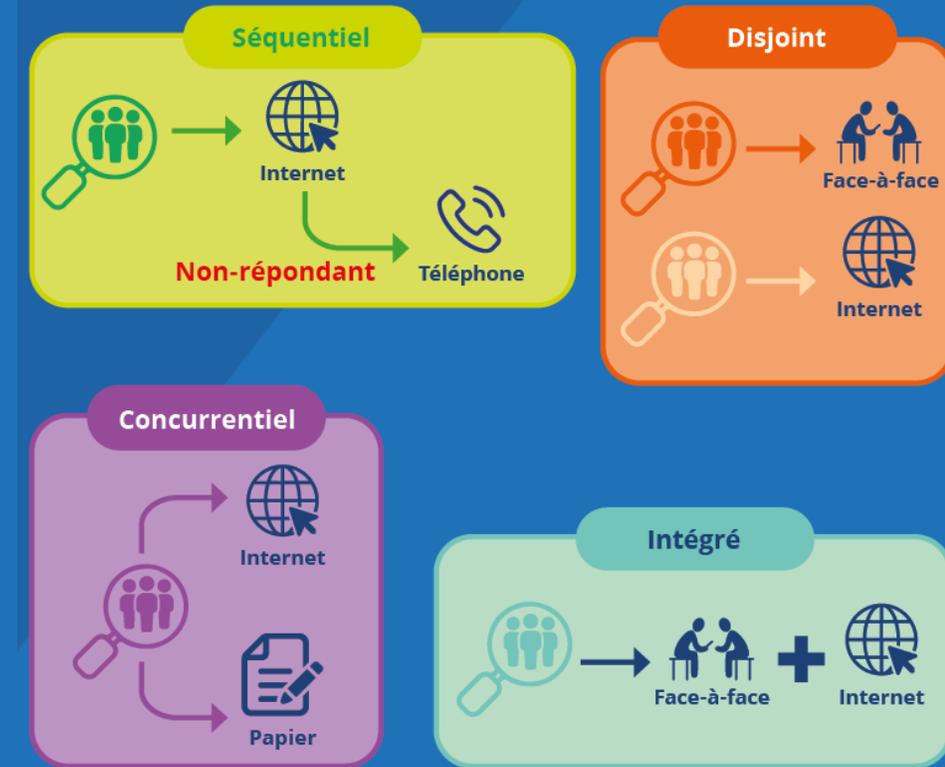


Tous les moyens sont bons, à condition de bien les combiner

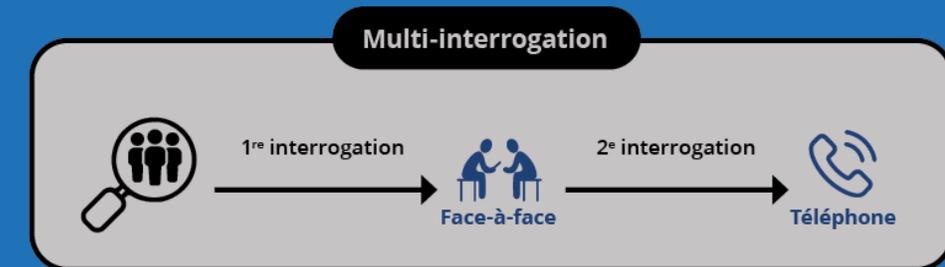
Source : F. Beck, L. Castell, S. Legleye, A. Schreiber, « [Le multimode dans les enquêtes auprès des ménages : une collecte modernisée, un processus complexifié](#) », *Courrier des statistiques*, 7, 2022, 7-28

	AVANTAGES	RISQUES OU INCONVÉNIENTS
 Internet	<ul style="list-style-type: none"> Coûts modérés Remontée d'informations rapide Possibilité d'introduire des filtres complexes Disponibilité des enquêtés Recours à des contenus variés (vidéos, son, images) Adapté aux 15-30 ans et aux populations connectées 	<ul style="list-style-type: none"> Non-réponse, auto-sélection Biais de <i>satisficing</i> Problèmes d'affichage Défaut de couverture Nécessite une bonne compréhension de la langue écrite Risque d'incompatibilité informatique
 Papier	<ul style="list-style-type: none"> Disponibilité des enquêtés Couverture large Adapté aux 75 ans et plus 	<ul style="list-style-type: none"> Remontée d'informations lente Aucune interactivité Impossibilité d'introduire des filtres complexes Non-réponse partielle Non-réponse, auto-sélection Nécessite une bonne compréhension de la langue écrite Biais de <i>satisficing</i>
 Téléphone	<ul style="list-style-type: none"> Intermédiation permettant un meilleur respect des consignes et des concepts Possibilité d'introduire des filtres complexes 	<ul style="list-style-type: none"> Pas de recours à des contenus variés (vidéos, son images) Défaut de couverture Peu adapté pour les 15-30 ans Biais de désirabilité sociale
 Face-à-face	<ul style="list-style-type: none"> Taux de réponse élevé Intermédiation permettant un meilleur respect des consignes et des concepts Possibilité d'introduire des filtres complexes Recours à des contenus variés (vidéos, son, images) Couverture large 	<ul style="list-style-type: none"> Coûts élevés Biais de désirabilité sociale Problème de disponibilité des enquêtés

Les différents modes (face-à-face, internet, téléphone, papier) peuvent se combiner pour des protocoles de collecte adaptés aux populations enquêtées...

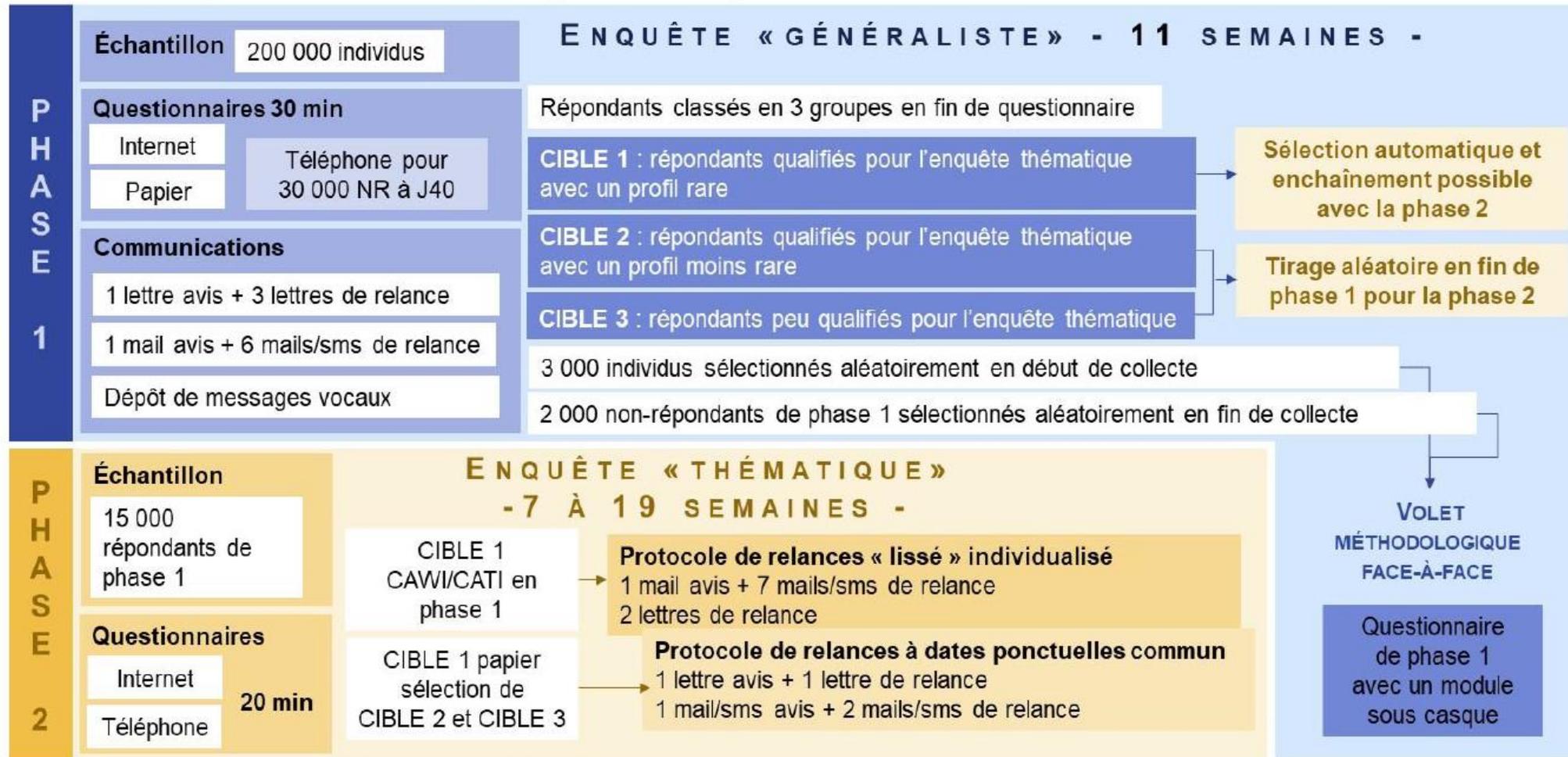


... et adaptés au rythme et aux thématiques des enquêtes



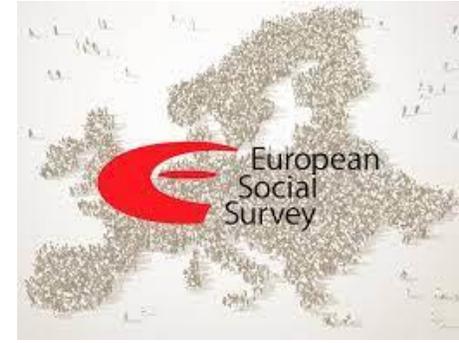
Nota : les combinaisons ci-dessus sont des exemples de protocoles multimode, ce ne sont pas les seules possibles.

L'exemple de l'enquête « Cadre de vie et sécurité » (CVS) (Insee en co-maîtrise d'ouvrage l'ONDRP et le SSMSI)



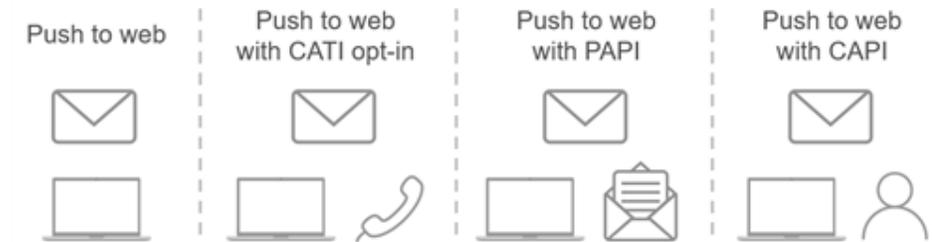
Source : H. Guedj, T. Razafindranovona et S. Zilloniz, « [Une enquête multimode annuelle pour mesurer la victimation en France à compter de 2022 : chronique d'une refonte en 4 actes](#) », *Statéco*, 116, 2022, 53-65

L'exemple de l'*European Social Survey*



- Triple constat :
 - Le choix simultané CAWI/PAPI n'améliore pas le taux de réponse (vs papier seul), mais le multimode séquentiel (CAWI puis PAPI) donne de meilleurs résultats.
 - Pour un multimode séquentiel CAWI+PAPI, retarder l'envoi ou la mention du questionnaire papier n'améliore pas le taux de réponse final, mais accroît la proportion des réponses online et limite donc les coûts.
 - On observe + de non réponses (partielles) online et - de variabilité sur les échelles de réponse, mais le biais de désirabilité sociale est - marqué online qu'en face-à-face
- Une expérimentation grandeur réelle (UK 2010) :
 - envoi postal d'un courrier avec lien web vers questionnaire à 8.000 addresses tirées aléatoirement
 - allocation aléatoire à un type d'incentive (inconditionnel) : aucun, £5, £10
 - 2 courriers de relance + relance finale avec questionnaire papier

- Un objectif : généraliser le modèle *push to web with PAPI* pour la vague 13 (2026)



3. Les nouveaux enjeux de la data

*Comment la Big Data challenge les
questions de collecte et d'échantillonnage*

Big Data et enquêtes quantitatives : V comme variété et volume



<https://www.youtube.com/watch?v=5otaBKsz7k4>

Les enjeux de la Big data (1/2)

- Liée au développement technologique (production, stockage, analyse), la big data implique le plus souvent une collecte passive (\neq données sollicitées et déclaratives)
 - Récolte des traces en ligne (réseaux sociaux, historique web)
 - Récolte des données de géolocalisation (via les mobiles notamment)
 - Récolte des données d'entreprises
 - Récolte des données liées à « l'internet des objets » et au « Web 3.0 » (via les objets connectés)
- La Big Data, ou « données massives » regroupe des données présentant une grande variété (de provenance et de nature), arrivant en volumes croissants, qui nécessitent une vitesse de traitement. C'est ce que l'on appelle les trois « V ».
- Prolifération des données. Si on sait les structurer et les analyser (création et l'application de modèles statistiques, d'algorithmes plus précis et plus riches), promesse d'une information plus facile d'accès, moins chère et permettant d'accéder à des informations non-couvertes par la donnée déclarative.

Les enjeux de la Big data (2/2)

- Mais d'autres enjeux connexes :
 - Véracité des données
 - Valeur (monétisation de la donnée) : quantité = qualité ?
 - Enjeux de privacy (pas de consentement)
 - Données incomplètes
- Menace ou opportunité pour la recherche ?
 - Immensité des données = multiplicité des acteurs. Fin du "monopole" des acteurs historiques (institutionnels ou non)
 - Données peu structurées = important travail méthodologique à effectuer
 - Les sciences sociales s'adaptent = complémentarité des données comportementales et déclaratives, open data (Data.gouv)...
- Pour aller plus loin, voir les travaux de Dominique Cardon, Julien Boyadjian ou Baptiste Kotras.

Big data et collecte des données par internet

- Nouvel écosystème/marché de la donnée
 - arrivée récente de nouveaux acteurs (access panels, mais aussi courtiers de données, plateformes de crowdsourcing et survey marketplaces)
 - généralisation des enquêtes online avec ciblage innovants des répondants (river sampling avec recrutement via des publicités, bannières, invitations... sur les réseaux sociaux)
- Le marché des panels a plutôt diversifié ses sources de recrutement (- incentive)
- Le contrat enquêteur-enquêté a changé, modèle plus pragmatique et transactionnel
 - L'autosélection (vs. devoir citoyen) est difficile à contourner et peut être un avantage
- Mais si l'on travaille avec des répondants issus de panels, important d'identifier :
 - les effets des conditions dans lesquelles les panélistes répondent (mode d'administration – effet device + temporalité)
 - Et surtout les conditions dans lesquelles les panélistes sont recrutés (canal, type et fréquence de sollicitation).



Quatre points à surveiller si la base de sondage a été recrutée sur internet

1. L'origine des fichiers d'adresses utilisés (par l'institut) : auprès de qui les fichiers ont-ils été achetés et combien y-a-il de niveaux de sous-traitance ? quelle est l'expérience des opérateurs ?
2. Le processus de sélection des individus recrutés : comment les répondants ont été sollicités pour répondre à l'enquête ? Est-ce que le message qui leur a été adressé fait directement mention du thème de l'enquête ? Est-ce qu'ils ont été sollicités par un ciblage préalable ou redirigés à partir d'un autre questionnaire ?
3. La fréquence de mobilisation des panélistes sur des études : combien de sollicitations ils reçoivent pour répondre à un sondage par semaine ? Jusqu'à combien de sondages est-il admis qu'ils participent ? Quelles sont les méthodes de vérification des profils et des comportements sont mises en œuvre ?
4. Le type de dédommagement reçu : permet de prendre une juste mesure du degré de (semi)professionnalisation des panélistes
 - Plusieurs risques d'inflation de l'erreur de mesure : un même panéliste peut développer des routines de réponses (straightlining), voire répondre plusieurs fois s'il est inscrit dans plusieurs listes (notamment s'il est une cible rare) ; risque de triche ; effets de priming et d'exposition préalable à un thème...
 - Les points 3 et 4 sont couverts dans la notice d'information déposée à la Commission des sondages. Ces questions sont aussi adressées par les organisations professionnelles (Esomar, [Questions to help buyers of online samples](#), mars 2021).