



Evaluation de la qualité de l'utilisation d'Internet dans une enquête de santé publique en population générale

Réplication de l'enquête aléatoire par téléphone FECOND au sein d'un panel d'internautes volontaires

Géraldine Charrance, Stéphane Legleye, Nicolas Razafindratsima
Service des enquêtes et des sondages, Ined



Plan de la présentation

- **Éléments de contexte**
- **Comparaison des collectes**
- **Validité des résultats obtenus dans des enquêtes online : Comparaison avec la même enquête menée par téléphone**
- **Éléments de conclusion**

Éléments de contexte

- Accès croissant des foyers à une connexion Internet
- Des avantages reconnus (collecte plus rapide, coûts plus faibles, pas d'effets enquêteurs, gestion facilitée des relances...)
- Distinguer Internet comme mode de collecte et univers d'échantillonnage
- De nombreuses études déjà réalisées (USA, NL, Royaume-Uni, etc.)
 - ✓ Échantillonnage
 - ✓ Taux de réponse
 - ✓ Vitesse de collecte
 - ✓ Qualité des données
 - ✓ Comparabilité des mesures avec d'autres modes
- Comparabilité et généralisation des résultats très largement inexplorées en France

Objectifs de l'étude

Comparaison d'une même enquête (FECOND) menée par téléphone et sur internet

- **Thème de l'enquête**

Etudier les différents enjeux en santé reproductive en France du point de vue des hommes et des femmes

- **Population cible**

Hommes et femmes âgés de 15 à 49 ans, résidant en France métropolitaine.

- **Taux de couverture plus faible pour Internet**

- Téléphonie fixe : 89 % / Téléphonie mobile : 84 % (99 % de la population dispose d'un téléphone)

- Internet (à domicile) : 75 % (population générale), 90% (15-50 ans)

- **Inégalités d'équipement au sein de la population**

- Internet semble encore aujourd'hui présenter d'importantes inégalités d'équipement et d'usages au sein de la population, contrairement au téléphone, très largement démocratisé

(Crédoc, La diffusion des technologies de l'information et de la communication dans la société française, 2011 ; Enquête Technologies de l'information et de la communication, Insee, 2011)

- **Echantillonnage**

→ **Téléphone & internet** : Pas de base de sondage existante

- **Méthode de constitution des échantillons FECOND**

→ **Téléphone**

- Sondage à deux degrés (génération aléatoire de numéros fixes et mobiles + sélection d'une personne dans le ménage)

→ **Internet**

- Tirage stratifié selon sexe, diplôme et âge d'un échantillon de répondants parmi un panel (marketing) de volontaires d'un institut de sondage

Collecte des données

○ **Durée de collecte et vitesse de rentrée des questionnaires**

→ Téléphone

- **210 jours** au total (41 questionnaires/jour)
- **145 jours** de collecte (60 questionnaires/jour)

→ Internet

- **34 jours** de collecte (265 questionnaires/jour)

○ **Techniques pour améliorer le taux de réponse**

→ Téléphone

- Nombre de tentatives d'appel élevé
- Rappel des refusants
- Incitation financière auprès d'une partie des refusants

→ Internet

- Envoi de plusieurs e-mails de relance, unique sollicitation d'enquête durant quinze jours

Taux de réponse, abandons

→ Téléphone : 8645 questionnaires retenus

- Taux de réponse : 44,8%
- 341 abandons, 8645 interviews : 3,8% des personnes ayant commencé le questionnaire ont abandonné en cours de route

→ Internet : 8992 questionnaires retenus

- Taux de réponse : 20,0%
- 1568 abandons, 9202 interviews : 14,6% des personnes ayant commencé le questionnaire ont abandonné en cours de route

Effets des techniques pour améliorer le taux de réponse



→ Téléphone

- Rappel des refusants : 19,0 % des questionnaires ont été recueillis après au moins un refus (1641 sur 8645)
- Nombreuses tentatives d'appel : 9,9% des questionnaires ont été recueillis après 20 appels minimum (859 sur 8645)

→ Internet

- Jusqu'à 4 emails de relance : 23,6% des questionnaires ont été recueillis après au moins une relance (2122 sur 8992)

Collecte des données



- **Durée de passation du questionnaire**

- **Téléphone**

- **41 minutes** en moyenne

- **Internet**

Difficulté de mesure car le questionnaire a pu rester ouvert sans que l'individu ne soit en train de le remplir.

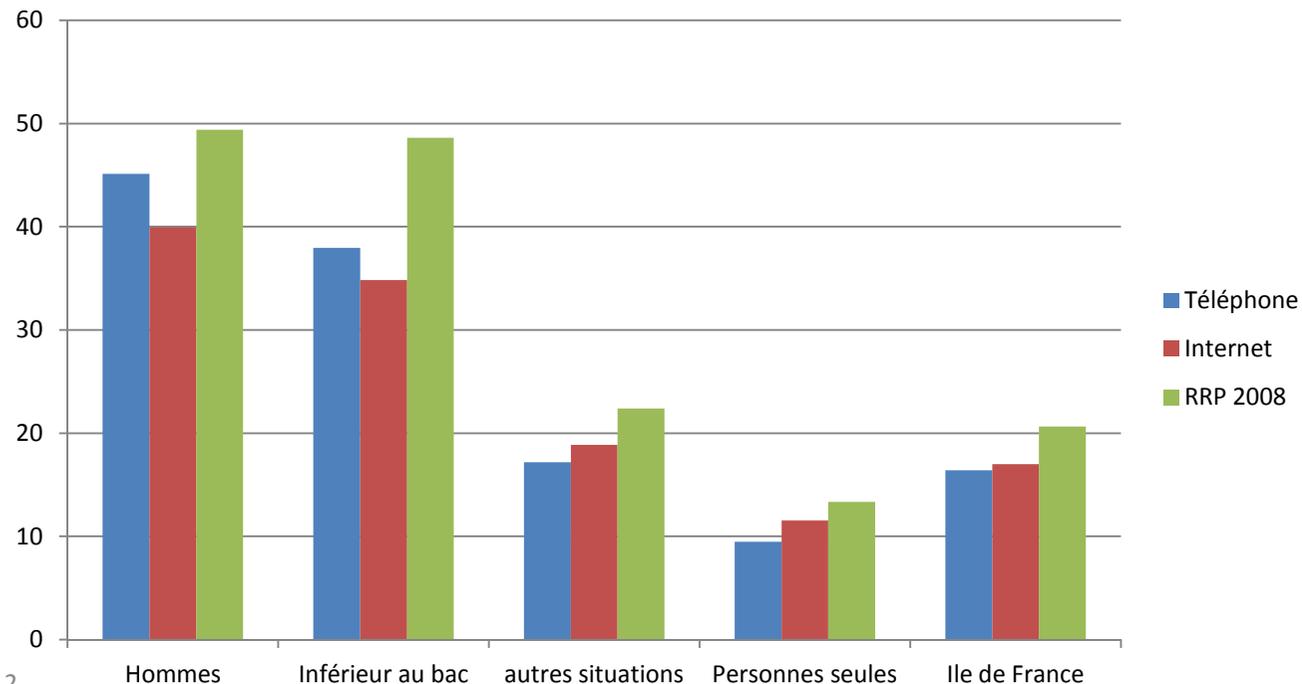
Pour estimer la durée de passation, deux options :

- Prendre uniquement en compte les interviews dont date de début = date de fin (87,7% des interviews) : **54 minutes** en moyenne
- Tronquer la durée de passation à 2h (83,6% des interviews): **40 minutes** en moyenne

Comparaison de la structure sociodémographique des échantillons avec la population cible

○ Populations sous-représentées dans les deux échantillons :

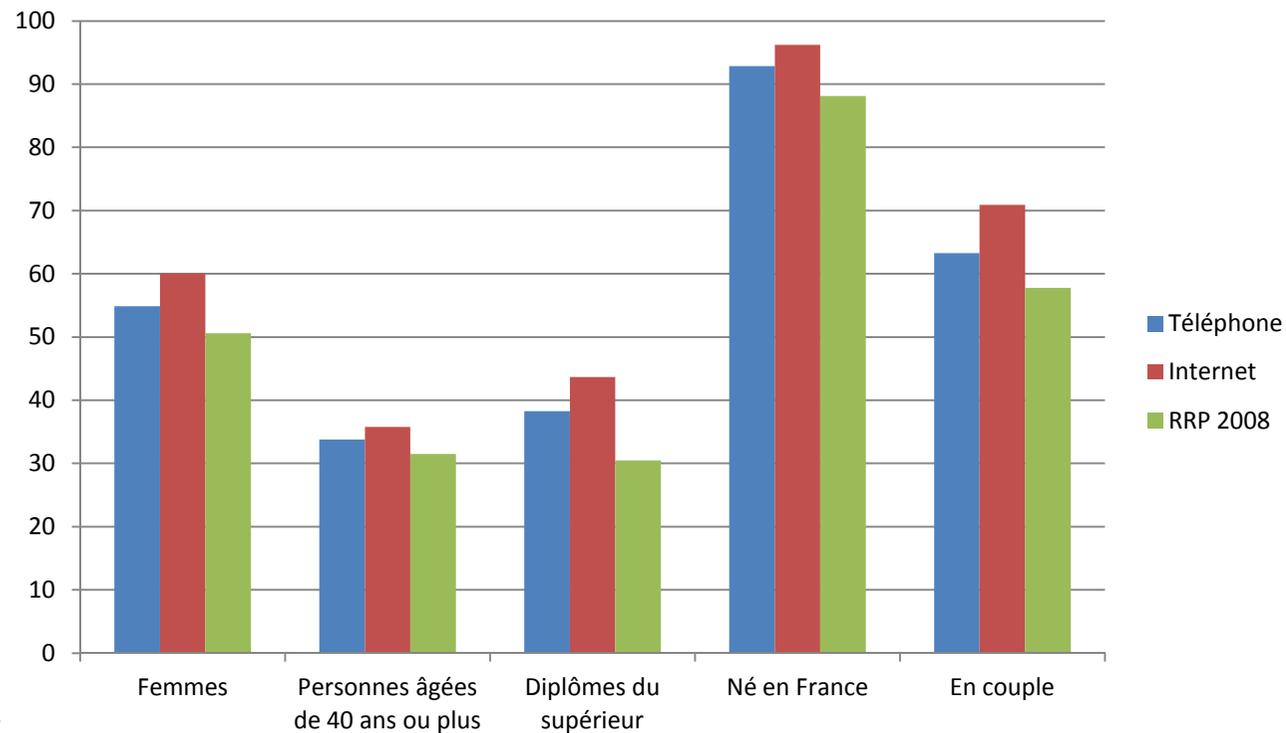
- Hommes
- Diplômes inférieurs au baccalauréat
- Personnes nées à l'étranger
- Personnes ne vivant ni en couple, ni chez leurs parents
- Personnes vivant seules
- Personnes résidant en Ile-de-France



Comparaison de la structure sociodémographique des échantillons avec la population cible

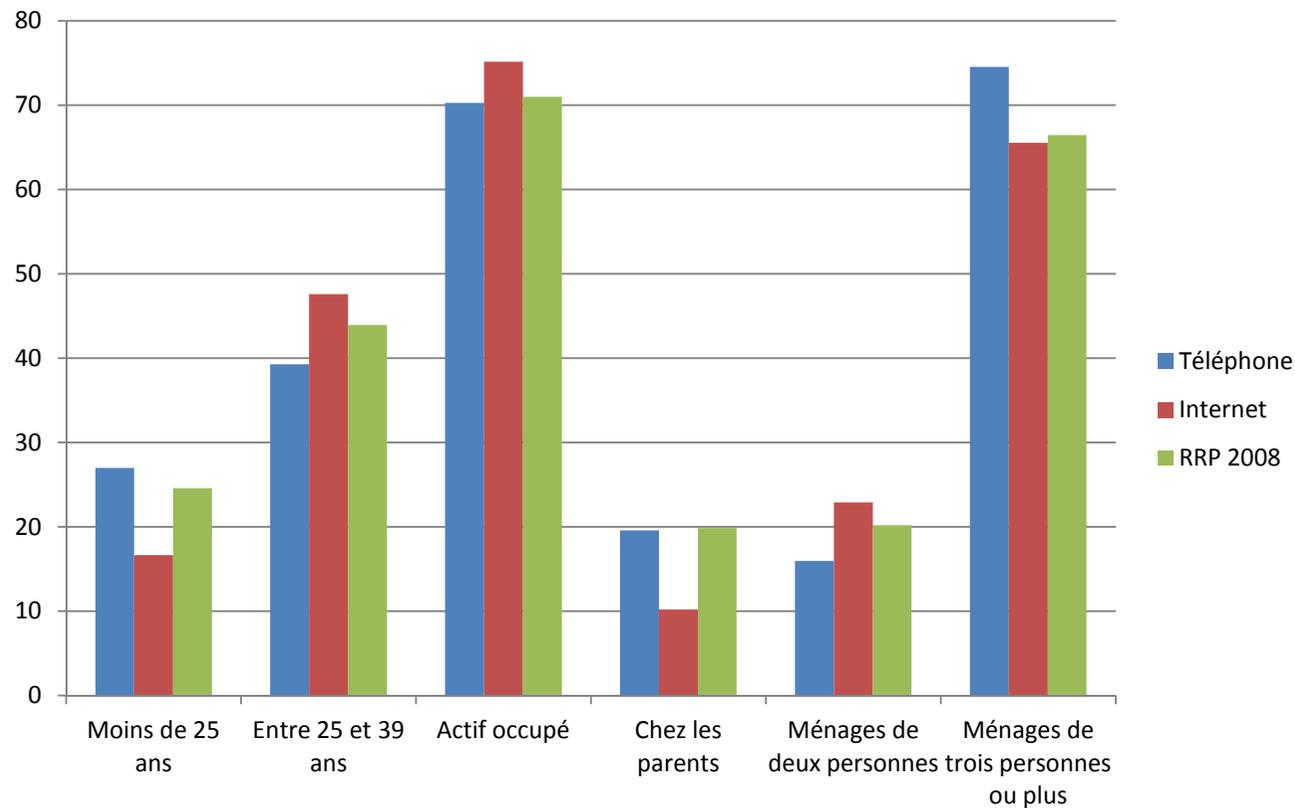
○ Populations surreprésentées dans les deux échantillons :

- Femmes
- Personnes de 40 ans ou plus
- Diplômes de l'enseignement supérieur
- Personnes nées en France
- Personnes vivant en couple



Comparaison de la structure sociodémographique des échantillons avec la population cible

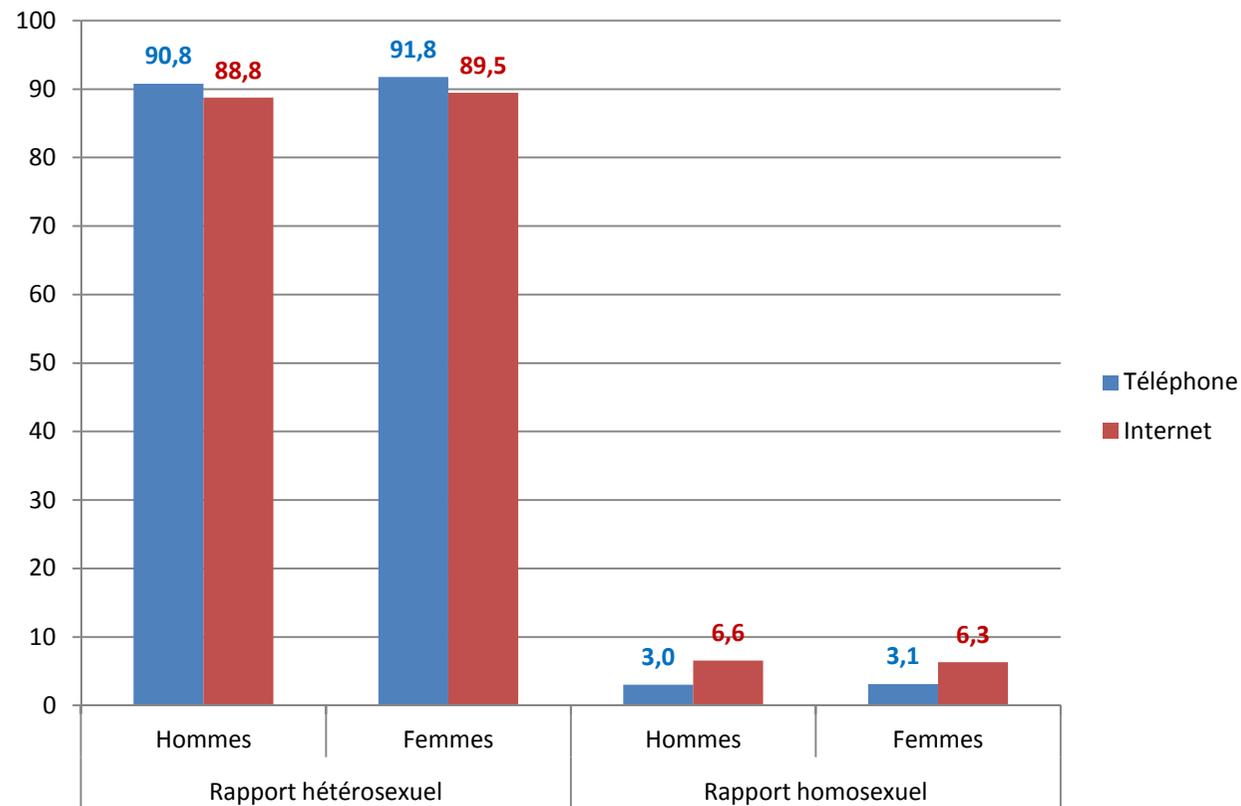
- Principales différences entre les deux échantillons :



Estimation des prévalences de quelques variables d'intérêt dans les deux échantillons (1)

→ Type de rapport sexuel au cours de la vie

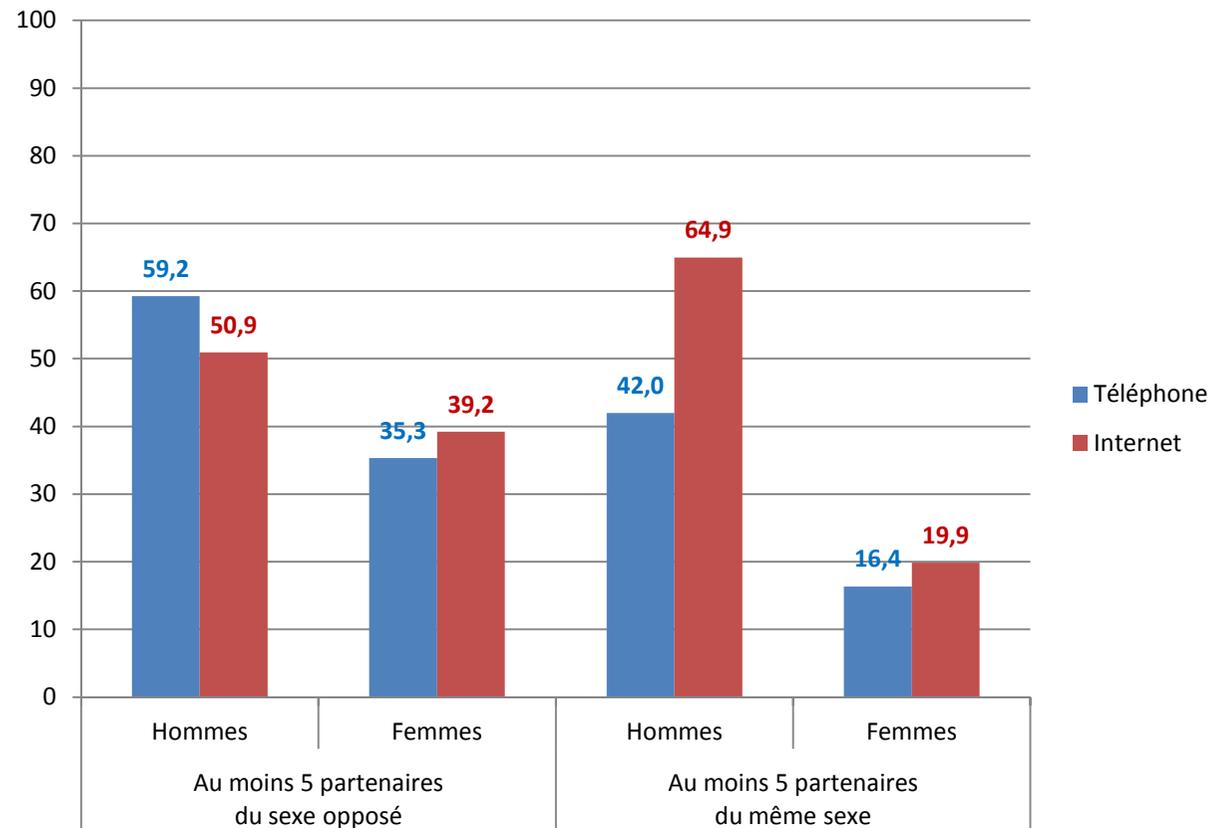
Différences significatives entre les échantillons téléphone et Internet



Estimation des prévalences de quelques variables d'intérêt dans les deux échantillons (2)

→ Nombre de partenaires sexuels au cours de la vie

Différences significatives et importantes entre les échantillons téléphone et Internet (hormis pour les partenaires du même sexe pour les femmes)

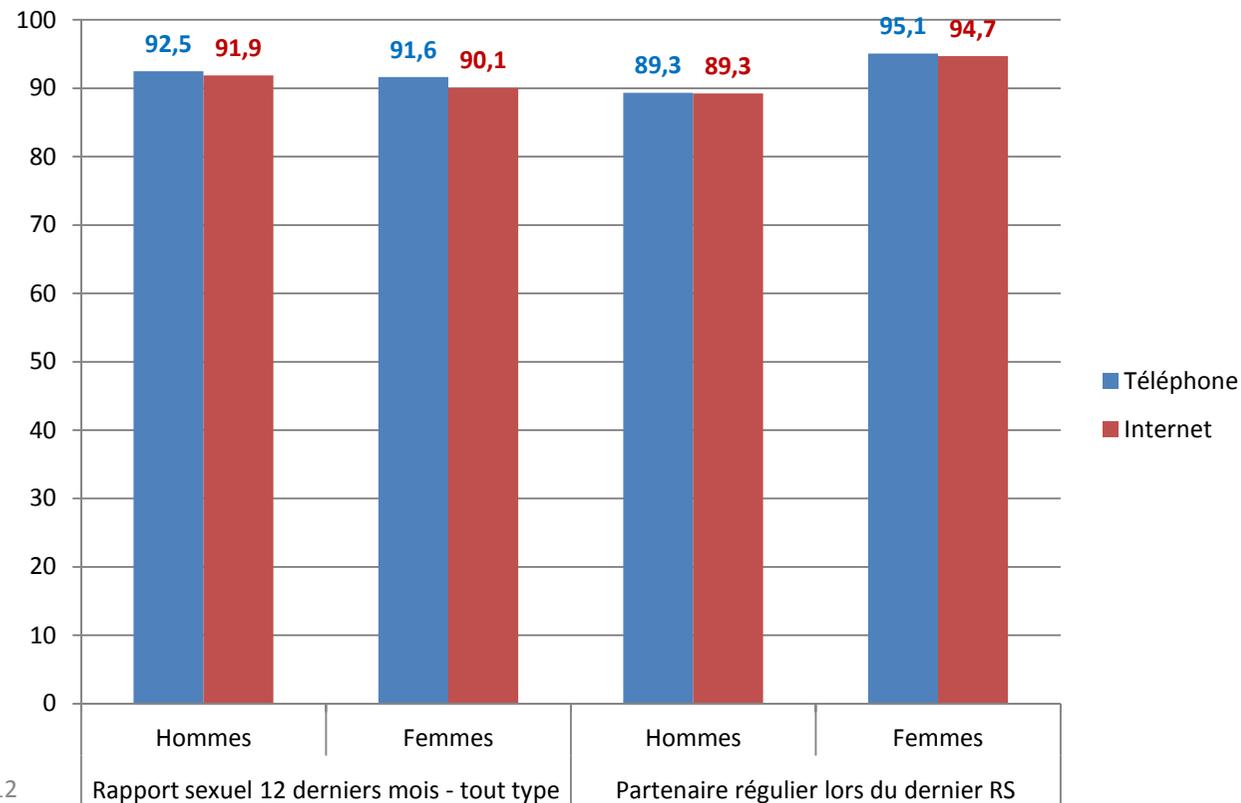


Estimation des prévalences de quelques variables d'intérêt dans les deux échantillons (3)



→ **Activité sexuelle récente**

Différence significative bien que faible entre les échantillons téléphone et Internet uniquement pour l'occurrence d'un rapport sexuel au cours des douze derniers mois chez les femmes



Comparaison des déterminants sociodémographiques de quelques comportements (1)



Modélisation avec interactions : Certaines caractéristiques jouent-elles différemment selon l'échantillon sur les comportements déclarés?

	Différence des déterminants	Caractéristiques sociodémographiques liées différemment à la variable d'intérêt dans les deux échantillons
Rapport hétérosexuel - Vie entière		
Hommes	✓ (seuil=5%)	Taille du ménage (5%) Mode de vie (10%)
Femmes	✗	Taille du ménage (5%)
Rapport homosexuel - Vie entière		
Hommes	✗	-
Femmes	✓ (seuil=10%)	Activité professionnelle (5%) Taille du ménage (10%)
Cinq partenaires sexuels du sexe opposé ou plus - Vie entière		
Hommes	✓ (seuil=10%)	Mode de collecte (5%)
Femmes	✓ (seuil=5%)	Diplôme (5%) Région (5%) Activité professionnelle (10%)

Comparaison des déterminants sociodémographiques de quelques comportements (2)



	Différence des déterminants	Caractéristiques sociodémographiques liées différemment à la variable d'intérêt dans les deux échantillons
Cinq partenaires sexuels du même sexe ou plus - Vie entière		
Hommes	X	-
Femmes	X	-
Au moins un rapport sexuel au cours des 12 derniers mois		
Hommes	✓ (seuil=5%)	Taille du ménage (5%)
Femmes	X	Mode de vie (5%) Région (10%)
Partenaire lors du dernier rapport sexuel : régulier/autre ?		
Hommes	✓ (seuil=10%)	Diplôme (5%) Mode de collecte (5%)
Femmes	X	Mode de vie (10%)

Bilan des résultats

→ La collecte par **Internet offre des avantages** en matière de :

- Coûts
- Rapidité
- Commodité de suivi du terrain
- Déontologie

→ **Prévalences (analyses bi-variées)**

Ecart variables, dépendant du sexe

- **Hommes** : différences significatives sur les variables relatives aux rapports sexuels vie entière (rapport hétérosexuel, rapport homosexuel, nombre de partenaires sexe opposé et même sexe – vie entière)
- **Femmes** : Différences sur l'occurrence d'un rapport hétérosexuel, rapport homosexuel, nombre de partenaires du sexe opposé et activité sexuelle des 12 derniers mois)

→ Déterminants des comportements sexuels (modèles logit)

- **Hommes**

Rejet de l'hypothèse d'égalité des coefficients pour toutes les variables sauf :

- occurrence d'un rapport homosexuel – vie entière
- Nombre de partenaires du même sexe – vie entière

- **Femmes**

Rejet de l'hypothèse d'égalité des coefficients pour :

- occurrence d'un rapport homosexuel – vie entière
- nombre de partenaires du sexe opposé – vie entière

L'âge et le lieu de naissance ne sont jamais liés différemment aux comportements étudiés sur les deux échantillons

→ L'étude des prévalences et des déterminants des variables d'intérêt a montré que **l'hypothèse de substituabilité des échantillons n'était pas vérifiée.**

Cependant, **intérêt exploratoire certain**

Robustesse pour les comportements communs

Différences notables pour l'homosexualité, mais pourrait aller dans le sens du biais connu de sous-déclaration au téléphone (effet enquêteur), en faveur d'Internet.

Recherches futures



- Etude sur d'autres variables d'intérêt (comportements dont les normes sont genrées)
- Effet des relances sur Internet : les apports sont-ils analogues à ceux observés pour le rappel de refusants ou bien l'inclusion de difficiles à joindre au téléphone?
- Différence de taux de réponse et de profils des répondants selon leur ancienneté dans le panel
- Articulation du téléphone et de l'Internet dans une collecte multi-modes : lancement d'un projet en ce sens en novembre 2012 (n=5000, 18-49 ans)
Permettra d'étudier Internet comme mode de collecte

Merci de votre attention

geraldine.charrance@ined.fr

Bibliographie (1)



D. de Leeuw Edith. 2012. « Counting and Measuring Online : The Quality of Internet Surveys », Bulletin de Méthodologie Sociologique 2012 114: 68

Hansen Kasper M., Pedersen Rasmus Tue. 2011. « Efficiency of different recruitment strategies for Web panels », International Journal of Public Opinion Research Advance Access published August 3, 2011

Nagelhout Gera E., Willemsen Marc C ., Thompson Mary E., Fong Geoffrey T., Putte Bas van den, Vries Hein de. 2010. « Is Web interviewing a good alternative to telephone interviewing ? Findings from the International Tobacco Control (ITC) Netherlands Survey », Public Health 2010, 10:351

Bowling Ann. 2005. « Mode of questionnaire administration can have serious effects on data quality », Journal of Public Health, vol. 27, No. 3, pp. 281-291

Spijkerman Renske, Knibbe Ronald, Knoops Kim, Mheen Dike van de, Eijnden Regina van den. 2009. « The utility of online panel surveys versus computer-assisted interviews in obtaining substance-use prevalence estimates in the Netherlands », *Addiction*, **104**, 1641–1645

Turner Charles F. et al.. 2009. « Improving epidemiological surveys of sexual behaviour conducted by telephone », International Journal of Epidemiology, 2009; 38:1118-1127

Bibliographie (2)



Vannieuwenhuyze Jorre, Loosveldt Geert, Molenberghs Geert, 2010. « A method for evaluating mode effects in mixed-mode surveys », Public opinion quarterly, vol. 74, No. 5, pp. 1027-1045

Schillewaert Niels, Meulemeester Pascale. 2005. « Comparing response distributions of offline and online data collection methods », International Journal of Market Research, vol. 47 issue 2

Braunsberger Karin, Wybenga Hans, Gates Roger, « A comparison of reliability between telephone and web-based surveys »

Ross Michael W., Mansson Sven-Axel, Daneback Kristian, Cooper Al, Tikkanen Ronny, 2005. « Biases in internet sexual health samples : Comparison of an internet sexuality survey and a national sexual health survey in Sweden », Social Science & Medicine, 61, pp. 245-252

Duffy Bobby and Smith Kate. 2005. « Comparing data from online and face-to-face surveys », International Journal of Market Research, vol. 47, issue 6

Dennis J. Michael, Li Rick. 2007. « More honest answers to web surveys ? A study of data collection mode effects », Journal of Online Research

Chang Linchiat, Krosnick Jon A., « Comparing oral interviewing with self-administrated computerized questionnaire », Public opinion quarterly, vol. 74, No. 1, Spring 2010, pp. 154-167

Bibliographie (3)



Chang Linchiat, Krosnick Jon A. 2009. « National Surveys via RDD telephone interviewing versus the Internet : Comparing sample representativeness and response quality ». *Public Opinion Quarterly*, vol. 73, No. 4, Winter 2009, pp. 641-678

Schonlau et al. 2003. « A comparison between responses from a propensity-weight web survey and an Identical RDD survey », *Social Science Computer Review*, Vol. 21 No. X, Season 2003 1-11

Dayan Yehuda, Schofield Paine Carina, Johnson AJ . «Responding to sensitive questions in surveys : A comparison of results from online panels, face to face and self-completion interviews ».

Hines Denise A., Douglas Emily M., Mahmood Sehar. 2010. « The effects of survey administration on disclosure rates to sensitive item among men : A comparison of an internet panel sample with a RDD telephone sample », *Comput Human Behav.* 2010 November 1; 26(6): 1327–1335.

Frippiat Didier et Marquis Nicolas, 2010, « Les enquêtes par Internet en sciences sociales : un état des lieux », *Population*, 2010/2, vol. 65, pp. 309-338