

# **Les défis posés par les sondages Internet**

Claire Durand,  
Professeur titulaire,  
Département de sociologie,  
Université de Montréal

© C. Durand, 2013

# Plan de la présentation

- Pourquoi parler de sondages Internet
- Qu'est-ce qu'un sondage Internet
- Problèmes spécifiques aux panels de volontaires
- Problèmes spécifiques à tous les sondages Internet
- Que faire?

# Pourquoi parler de sondages Internet?

- Prolifération de sondages électoraux ou socio-politiques utilisant des panels Internet
- L'accès à Internet se démocratise et l'accès sans fil, particulièrement dans les pays moins développés, est susceptible d'augmenter rapidement.
  - Pénétration: 14% Algérie, 33% Maroc et Tunisie.
- Internet est très économique, ce qui favorise le recours à ce mode y compris dans les pays moins développés.
- Des sondages "internationaux" présentent une "image" de plusieurs pays d'Afrique, d'Asie basée sur des sondages Internet.

# Qu'est-ce qu'un sondage Internet

- Bien sûr, une enquête via site web mais, quel mode de recrutement? Couper (2000) présente 8 modes de recrutement différents:
  - ▶ Non probabilistes:
    - 1. Vox pop sur les sites des medias
    - 2. Recrutement de volontaires pour 1 enquête (\$)
    - 3. Recrutement de volontaires pour panel (\$)
  - ▶ Probabilistes:
    - 4. Sondages de sortie de site
    - 5. Listes courriel complètes (organisations)
    - 6. Sondages multi-modes avec possibilité d'utiliser Internet
    - 7. Panels d'utilisateurs d'Internet recrutés par un autre mode.
    - 8. Échantillons probabilistes de l'ensemble de la population où l'on fournit l'accès Web lorsque le ménage ne l'a pas.

# Qu'est-ce qu'un sondage Internet?

D'où il ressort que...

- En ce moment, seul le mode 8, où on recrute par un autre moyen et où on fournit l'accès Web aux ménages non équipés, permet de constituer un échantillon probabiliste de la population générale.
- Or, pour le moment, ce mode est très cher et peu disponible, ce qui enlève une partie de l'intérêt d'utiliser Internet.
- La plupart des sondages Internet visant l'ensemble de la population utilise un échantillon non probabiliste d'internautes inscrits à un panel (mode 3).

# Quels problèmes pour les panels Internet de volontaires?

Les problèmes de couverture et d'auto-sélection

- Lorsque l'on compare les enquêtes de type panel de volontaires et les enquêtes population faites par un autre mode, deux problèmes d'entremêlent:
  - ▶ Un problème de couverture: Une partie de la population, plus ou moins importante selon les pays, n'a pas accès à Internet. Cette fraction de la population a des caractéristiques distinctives.
  - ▶ Un problème de recrutement et d'auto-sélection: Les modes de sollicitation à devenir membre de panels varient et ne rejoignent pas tous les internautes, ni ne les sélectionnent. Par ailleurs, pour la plupart des panels, les répondants peuvent décider de s'inscrire en allant sur le site de l'institut de sondage.

# Quels problèmes pour les panels Internet de volontaires?

## Quelques études

- Bigot, Croutte et Recours (2010) comparent simplement des répondants selon leur accès à Internet.
- Blasius et Brandt (2010) ont tenté de constituer un échantillon “représentatif” (par quotas) de la population à partir des panelistes d’un sondage Internet.
- Loosveldt et Sonck (2008) regardent si la pondération peut rendre les résultats d’un sondage panel Internet similaires à ceux d’un sondage face à face.
- Malhotra et Krosnick (2007) et Pasek et Krosnik (2010) comparent également des panels Internet et des sondages face à face.
- Stephenson et Crête (2011) comparent un panel Web et un sondage téléphonique de type GANT.

# Bigot, Croutte et Recours (2010)

- Différences entre internautes et non-internautes en France:
  - ▶ Non redressé: différences significatives pour 71 variables sur 191 (34%) portant plus particulièrement sur
    - L'ensemble des variables socio-démographiques à l'exception du genre,
    - L'équipement des ménages mais également
    - Les opinions en matière de moeurs
  - ▶ Après redressement: différences sur 12% des variables dont
    - L'équipement des ménages
    - Les opinions en matière de moeurs
    - Le logement
    - Les pratiques culturelles



# Blasius et Brandt (2010)

- Comparent un échantillon de panel représentatif des 18-49 ans à des échantillons “étalon” en face à face (GSS et Micro-recensement) en Allemagne.
  - Ils constatent qu’il est impossible d’avoir suffisamment de répondants de 50 ans et plus.
  - Ils réussissent à faire un échantillon des 18 à 49 ans représentant cette population de façon proportionnelle selon l’âge, le sexe et l’éducation.
- Les comparaisons (y compris après pondération) avec les 2 autres échantillons montrent, dans le sondage Internet:
  - Plus de célibataires ou divorcés, de personnes sans enfants
  - Moins de personnes fréquentant l’église.
  - Plus de valorisation du laisser faire, de la richesse, de la réalisation de soi.

# Loosveldt et Sonck (2008)

- Comparent un sondage Internet de type panel et un sondage face à face en Région flamande (Belgique).
- Regardent d'abord l'accès Internet pour pouvoir corriger par un score de propension (similaire à Bigot et coll.).
  - ▶ 18% des personnes ayant moins de 10 ans de scolarité ou ayant 60 ans et plus ont accès à Internet.
  - ▶ Comparé à plus de 80% des personnes en emploi, ayant une scolarité universitaire, ayant 30 ans ou moins.
- La pondération par score de propension permet d'ajuster pour les différences dans la proportion d'urbains et de personnes en emploi.
  - ▶ Mais les différences demeurent significatives pour ce qui est de la satisfaction face à l'emploi (Internet - ), de l'intérêt pour la politique (+) et des attitudes face aux immigrants (plus négatives chez les répondants au panel Internet).

# Malhotra & Krosnick, 2007

- Comparent l'American National Election Study de 2000 et celui de 2004 à des sondages Internet de type panel de volontaires (U.S.A.).
- Ils montrent que:
  - Même après pondération, les sondages internet comportent:
    - Moins de Noirs, moins de personnes peu scolarisées et près de deux fois plus de personnes ayant une scolarité moyenne.
    - En 2004, plus de personnes ayant voté, plus de partisans de Bush (par rapport à Kerry), plus de partisans de la guerre en Irak, plus de personnes intéressées par la politique.
    - En 2000, à peu près les mêmes différences qu'en 2004. À peu près deux fois plus de "strong Republicans".
    - De plus, les relations entre les variables prédisant le vote et le vote lui-même étaient significativement différentes dans presque tous les cas.

# Pasek et Krosnick (2010)

- On aurait pu penser que la situation s'améliorerait avec le temps mais...
- Pasek et Krosnick comparent un sondage téléphonique de type RDD et un panel Internet de volontaires sur la prédiction de l'intention de collaborer au recensement et la collaboration effective. Les sondages Internet de type panel
  - ▶ Ont une moins bonne distribution démographique
  - ▶ Diffèrent en moyenne de 13 points et vont jusqu'à 30 points dans les proportions des réponses modales.
  - ▶ Présentent des différences significatives et non insignifiantes dans la prédiction de la participation au recensement de même que l'évolution dans le temps des opinions et les relations entre les variables.

# Stephenson et Crête (2011)

- Comparent deux enquêtes faites par Léger Marketing, une par panel Web, l'autre par sondage téléphonique probabiliste.
- Il y a des différences significatives entre les deux échantillons
  - ▶ avant pondération pour 40 variables sur 52
  - ▶ après pondération, pour 36 variables sur 52
  - ▶ Les différences varient de 2 à 28 points de pourcentage.
  - ▶ Pour les relations entre 3 des 23 variables liées aux déterminants du vote.
- Membres du panel sont
  - ▶ Plus cyniques face à la politique
  - ▶ Plus intéressés à la politique, moins aux élections
  - ▶ Plus de partisans du Parti vert, moins du PQ

# En résumé, des biais relatifs à la couverture et au mode de recrutement

- Différences fréquentes entre les sondages Internet de type panel de volontaires et les méthodes probabilistes en face à face ou au téléphone au désavantage des sondages Internet.
  - Des différences non constantes, non systématiques et donc non prévisibles et difficiles à corriger.
- Lorsque des quotas sont appliqués avec succès, on constate des problèmes similaires à ceux d'autres échantillons par quotas: le respect des quotas n'assure pas nécessairement une représentation socio-politique adéquate.
- On estime que 3% des internautes complètent plus de 80% des sondages Internet (Rivers, Yougov).

# Au delà des problèmes spécifiques aux panels de volontaires...

## Les taux de réponse

- Lozar Manfreda et coll., 2008 font une méta-analyse de 45 expériences comparant le mode Internet à d'autres modes.
  - Ils montrent que les taux de réponse aux sondages Internet sont généralement plus bas de 11% en moyenne (6%-15%) que ceux des autres modes.
- Pour les **sondages de type panel** de volontaires, on parle en ce moment de taux aussi bas que moins de 1%, ce qui empire probablement le problème de représentation et amène les firmes à essayer de garder les membres de leurs panels plus longtemps.
- **Pour les autres sondages, pour atteindre des taux de réponse similaires aux sondages auto-administrés "classiques", il faut une gestion très serrée, des rappels, etc.**

# Au delà des problèmes spécifiques aux panels de volontaires...

- Le questionnaire:
  - ▶ Dans une entrevue, l'interviewer peut "aider" le répondant au besoin, expliquer.
  - ▶ Dans un questionnaire en format papier, le répondant voit l'instrument au complet, les questions qui viennent avant et après, leur nombre, leur forme.
- Dans un sondage sur Internet, les questions "disparaissent" aussitôt répondues, ce qui rend la tâche du répondant plus difficile. Il faut tenir compte de cette situation:
  - ▶ Donner des indications sur la progression du questionnaire.
  - ▶ Grouper les variables en format tableau lorsque le choix de réponse est le même pour une série d'énoncés.
  - ▶ Insérer des phrases de transition lorsqu'on fait sauter des questions pour permettre au répondant de mieux se situer.
  - ▶ Permettre la non réponse sauf lorsque la réponse est essentielle et la question non sensible pour éviter les réponses "au hasard".
  - ▶ Permettre le retour en arrière.



# Au-delà des problèmes spécifiques...

## La confidentialité

- Les répondants ont-ils autant confiance dans les assurances relatives à la confidentialité des réponses dans un sondage Internet? Comment les rassurer? (Manfreda Lozar et coll., 2008)
- Quelles informations doit-on recueillir sur les répondants?
  - ▶ Sans doute pour prévenir les inscriptions multiples, certains instituts recueillent des informations tellement détaillées, personnelles -- date de naissance, adresse exacte, numéro de téléphone en plus de l'adresse courriel -- qu'on peut se demander si cela n'entraîne pas un biais supplémentaire d'auto-sélection.

# Pourquoi donc fait-on des sondages Internet?

- Certainement pour une question de coûts.
- Si on a accès à une liste d'adresses courriel d'une population spécifique (employés d'une entreprise, membres d'une organisation, étudiants d'une université), le sondage Internet est sans doute **le** meilleur moyen de consulter ces populations.
  - ▶ Possibilité d'avoir recours à des présentations visuelles.
  - ▶ Recours aux questions ouvertes et analyse des réponses à celles-ci facilités.
  - ▶ Mais il faut prendre les moyens pour créer un très bon instrument et obtenir un bon taux de réponse.
- Pour les sondages visant la population générale, en l'absence de base de sondage fiable en ce moment, les coûts pour faire un sondage avec échantillon probabiliste sont prohibitifs.

# Les sondages Internet, présent et avenir

## ■ Présent:

- ▶ Les sondages internet de populations spécifiques qui peuvent être entièrement rejointes par courriel sont des outils tout à fait appropriés.
- ▶ Les sondages de type panel de volontaires se font une place essentiellement pour des questions de coût et dû au manque d'information et de formation de la population et des journalistes.

## ■ Avenir:

- ▶ Les sondages Internet sont l'avenir mais... on est dans le présent.
- ▶ La population mondiale sera éventuellement couverte.
- ▶ Les outils permettant de constituer des échantillons probabilistes de la population commencent à se développer.

# En conclusion

- Entretiens, formation et information sont essentielles si on ne veut pas que tôt ou tard des sondages biaisés influencent indûment les décisions politiques ou même le vote.