

# Enquêtes téléphoniques auprès des ménages

la génération aléatoire de numéro

Stéphane Legleye (Insee/Inserm)

Noémie Soullier (Santé publique France)

Jean-Baptiste Richard (Santé publique France, Inserm)

# I. Les enquêtes téléphoniques

les avantages du téléphone comme mode de collecte (1)

- Pas de nécessité de maîtriser l'écrit, les technologies de l'information
- Bon marché (relativement au face-à-face)
- Suivi et contrôle de la collecte
  - Si collecte centralisée :
    - Écoutes et soutien en direct
    - Débriefings faciles
  - Recueil informatisé de parodonnées : heures et jours d'appel, durée de la collecte, historique des tentatives etc.

# I. Les enquêtes téléphoniques

les avantages du téléphone comme mode de collecte (2)

- Qualité de la collecte grâce aux enquêteurs
  - Reformulations
  - Soutien, motivation etc.
  - Possibilité de questionnaires relativement longs : peu de satisficing

MAIS:

- Désirabilité sociale pour les thématiques sensibles
- Durée de questionnaire acceptable inférieure à celle du face-à-face (25-30 mn en France)

# I. Les enquêtes téléphoniques

## Sondage et téléphone: le cas des bases de sondage (1)

- Base de sondage riche et exhaustive
  - Exemple: annuaire des abonnés
- Définition relativement claire
  - Du lien entre un numéro et une unité statistique cible (ménage, individu)
  - d'une réponse (participation) ; d'une non-réponse : d'un refus, d'une impossibilité à joindre, d'une ligne téléphonique en dérangement etc.
  - Des hors champs (entreprises)
- Accès à des informations auxiliaires
  - Type et localisation de l'abonnement, caractéristiques de l'abonné (éventuellement)
- Possibilité d'envoyer une lettre-annonce

# I. Les enquêtes téléphoniques

## Sondage et téléphone: le cas des bases de sondage (2)

- En pratique, il peut y avoir une ou deux bases :
    - Numéros filaires: liaison physique à un ménage, une adresse
      - Lien entre un ménage (le plus souvent) et un numéro
    - Numéros mobiles: liaison logique avec un individu
- ➔ La base de sondage complète est donc bisource (*overlapping dual frame*)  
Ces deux sources ne sont pas indépendantes et il faut en tenir compte!

# I. Les enquêtes téléphoniques

## Sondage et téléphone: le cas des bases de sondage (3)

- En pratique, les bases peuvent être incomplètes :
  - Par exemple, possibilité pour les abonnés de ne pas apparaître dans l'annuaire
- ➔ La base de sondage complète est donc bisource (*overlapping dual frame*) et incomplète

# I. Les enquêtes téléphoniques

## Le cas du Baromètre santé (1)

- Jusqu'en 2005 :
  - **Fort taux d'équipement filaire (96%), annuaire national centralisé filaire**
    - 10% de numéros en « listes rouges » (ménages hors annuaire)
  - Peu de téléphones mobiles, très peu d'abonnés mobiles sans numéro filaire
  - → Plus de 95% de la population joignable via un téléphone (filaire)
- → Pas de prise en compte des téléphones mobiles
- Tirages de numéros filaires dans l'annuaire (sondages stratifiés régionaux possibles)
- Ajout de +1 pour obtenir les « listes rouges »
- Usage de l'annuaire inversé pour repérer les numéros dans le champ et hors champ
- Interview

# I. Les enquêtes téléphoniques

## Le cas du Baromètre santé (2)

- 2005- : génération aléatoire (*random digital dialing*: RDD)
  - Annuaire plus exhaustif; perte du lien géographique (« dégroupage »)
- 2005-2014: Echantillonnage RDD bisource disjoint (*disjoint dual frame*)
  - → Génération aléatoire de numéros fixes
    - Usage de l'annuaire inversé pour repérer les numéros éligibles
  - → Génération aléatoire de numéros mobiles
    - On garde uniquement les **individus ne disposant pas d'un numéro fixe (filtre durant l'interview)**
- 2014- : RDD *overlapping dual frames*
  - Génération aléatoires de numéros fixes et mobiles
  - Prise en compte de l'usage de l'équipement pour définir la pseudo-probabilité de tirage



# Rappel sur les sondages probabilistes

- Connaissance de la qualité de la base de sondage :
  - Unicité des unités de la base
  - Lien unité de la base – unité cible (ménage, individu)
  - Fraîcheur et exhaustivité de la base
- Assignation d'une probabilité non-nulle et connue à toutes les unités statistiques de la base de sondage
- Connaissance du lien entre l'unité échantillonnée et l'unité statistique interrogée
  - Par exemple :
    - adresse postale – ménage ; ménage – individu
    - Numéro de téléphone – ménage ; ménage – individu

# En l'absence de base... (1)

- Assomption de la joignabilité de la population via le téléphone
  - Couverture théorique : taux de ménages / individus équipés
  - Couverture pratique: tout le monde ne décroche pas
  - Connaître cette couverture (par d'autres enquêtes) est essentiel
  - Si la couverture réelle est bonne, l'enquête pourra être « représentative » (éventuellement...)
- Il faut
  - constituer une pseudo-base représentative (génération aléatoire)
  - **reconstituer le lien entre l'unité de la base et l'unité statistique**

# En l'absence de base... (2)

- En pratique:
  - Génération de numéros aléatoires
  - Premier filtrage: appariement à la base nationale des numéros ouverts à l'exploitation commerciale (ARCEP)
    - Par exemple :
      - les numéros en 08 ne sont pas des numéros pour les particuliers
      - La tranche « 01 02 03 nn nn nn » n'est pas ouverte à l'exploitation
  - Deuxième filtrage: annuaire des entreprises (si disponible)
  - Troisième filtrage: éligibilité du ménage / individu répondant
- Pas de stratification géographique; pas de lettre-annonce

# En l'absence de base... (2)

- Ajouter des questions relatives à l'équipement et à l'usage du téléphone dans le questionnaire
- Objectif: connaître le nombre de lignes utiles à l'enquête
- Deux séries de questions:
  - Nombre de numéros fixes et mobiles permettant de joindre l'unité statistique (ménage, individu)
  - Pratique de décroche : sur combien de lignes téléphoniques fixes et mobiles l'unité statistique ne décroche jamais lorsqu'elle ne connaît pas le numéro?

# Les questions supplémentaires (1)

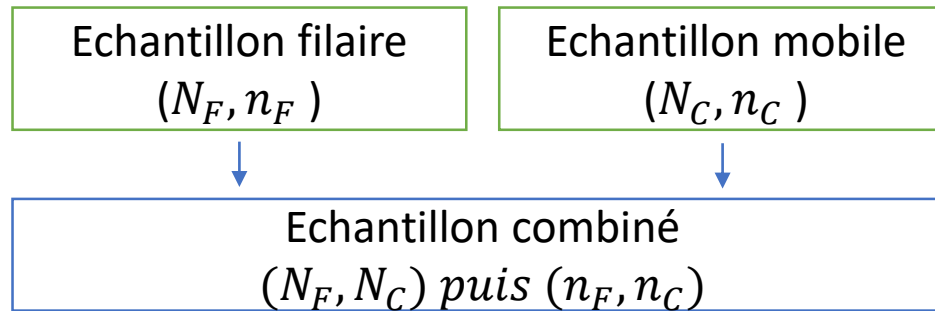
- **De combien de numéros de téléphone fixe disposez-vous dans votre ménage (numéros commençant par 01 à 05 ou par 09) ?**
  - Aucun
  - Un numéro
  - Deux numéros
  - Trois numéros
  
- **Décrochez-vous ou une personne de votre ménage décroche-t-elle lorsque votre téléphone fixe sonne ?**
  - Oui, mais seulement lorsque vous connaissez le numéro qui essaie de vous joindre
  - Oui, sans vous soucier du numéro qui essaie de vous joindre
  - Non, jamais
  - Non, il ne sonne jamais

# Les questions supplémentaires (2)

- **Pour votre usage privé, utilisez-vous ? (plusieurs choix possibles)**
  - Un smartphone
  - Un téléphone mobile hors smartphone
  - Aucun de ces appareils
- **Décrochez-vous lorsque votre téléphone mobile sonne ?**
  - Oui, mais seulement lorsque vous connaissez le numéro ou le nom de la personne qui essaie de vous joindre
  - Oui, sans vous soucier du numéro qui essaie de vous joindre
  - Non, jamais
  - Non, il ne sonne jamais

# Calcul des pseudo-poids de sondage

$e_F^i$  = nb d'individus éligibles avec le numéro filaire  $i$   
 $t_F^i$  = nb de numéros filaires de  $i$   
 $n_F$  = nb de répondants filaires  
 $N_F$  = Nb de numéros filaires tirés



$e_C^i$  = nb d'individus éligibles avec le numéro cellulaire  $i$  (=1)  
 $t_C^i$  = nb de téléphones cellulaires de  $i$   
 $n_C$  = nb de répondants cellulaires  
 $N_C$  = nb de numéros cellulaires tirés

$$\pi_i = \frac{n_F}{N_F} * \frac{t_F^i}{e_F^i} + \frac{n_C}{N_C} * \frac{t_C^i}{e_C^i}$$

Calage

# Quelques remarques critiques

- Il vaut mieux disposer de la proportion de ménages/individus équipés de lignes fixes et mobiles dans la population pour décider la part de numéros mobiles à générer (nombres de  $N_{LL}$  et  $N_c$ )
- Si possible, connaître en amont la proportion d'abonnements d'entreprises (hors champ); la ventilation des abonnements fixes et mobiles par critères sociodémographiques et géographiques
- **Pas de correction de la non-réponse totale avant le calage** : on ne dispose d'aucune information sur les non-répondants
  - le redressement n'est qu'un calage direct  
(On peut inclure des variables de post-stratification géographique dans le calage si des données démographiques régionales sont disponibles)



# Une difficulté notoire: calculer les taux de réponse

- Base de sondage : lien connu entre l'adresse ou le moyen de contact et l'unité statistique
  - Sur les répondants et les non-répondants
- Génération aléatoire: on ne connaît pas à l'avance le nombre de numéros donnant accès à un individu
  - Des numéros sont ouverts à l'exploitation mais utilisés par des machines
  - Des numéros de « boxes » Internet sans combiné téléphonique en service
  - Aucun moyen de savoir
  - Utiliser la grille de calcul de l'AAPOR

# Merci pour votre attention

## Références:

- Richard JB, Andler R, Guignard R, et al. Baromètre santé 2017. Méthode d'enquête. Objectifs, contexte de mise en place et protocole. Saint-Maurice: Santé publique France, 2018. p. 24.  
<http://inpes.santepubliquefrance.fr/CFESBases/catalogue/pdf/1856.pdf>