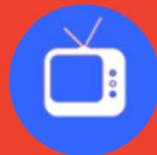


# Les sondages médias : évolutions et perspectives



7<sup>ème</sup> Colloque francophone sur les  
sondages

Rennes, le 6 novembre 2012



## **La fin du principe d'unicité**

- ▶ Encore au début des années 1990 le principe de base était simple : un questionnaire uniforme administré selon un unique mode de recueil de l'information
- ▶ Les innovations technologiques ont changé ce paradigme d'unicité, nous conduisant à faire évoluer les méthodologies et à aller vers des systèmes d'observation de plus en plus complexes
- ▶ L'objet de cette communication est d'évoquer les grandes étapes de cette mutation



# Plan de la présentation

## **Les enquêtes multi-canal**

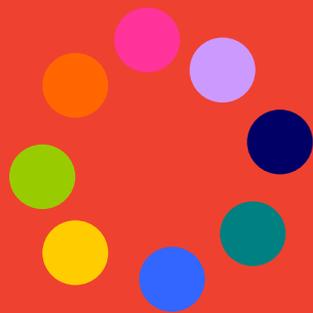
- ▶ Constats
- ▶ Nouveaux protocoles de recueil
- ▶ Conséquences sur l'échantillon

## **Les mesures hybrides**

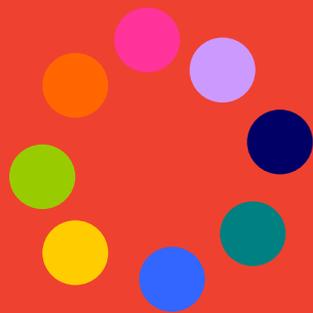
- ▶ Définition
- ▶ Méthodes proposées

## **Perspectives et enjeux en matière de R&D**

# Les enquêtes multi-canal



# *Les constats*





# Les difficultés actuelles

## Joignabilité dans les enquêtes téléphoniques

- ▶ En moyenne, 11 numéros et 48 appels sont nécessaires pour réaliser une interview dans l'enquête 126 000 Radio en 2012, avec une tendance à l'augmentation
- ▶ Un nombre de refus stable au global, mais un ressenti de l'augmentation des refus virulents depuis 2010
- ▶ Des cibles particulièrement difficiles à joindre

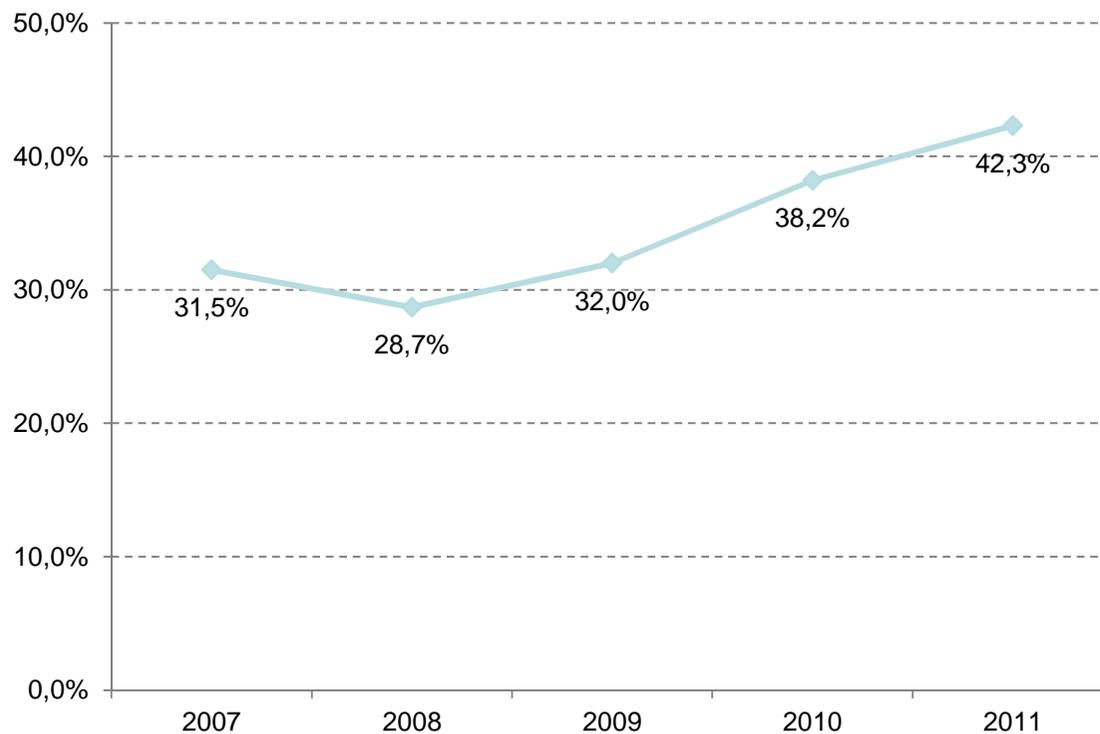
## Participation aux panels

- ▶ 1,7 fois plus d'adresses nécessaires en 2012 qu'en 2002 pour recruter le même nombre de panélistes pour le panel Radio
- ▶ Des cibles dont le taux d'abandon s'accroît d'année en année



# Augmentation du taux de sans réponse

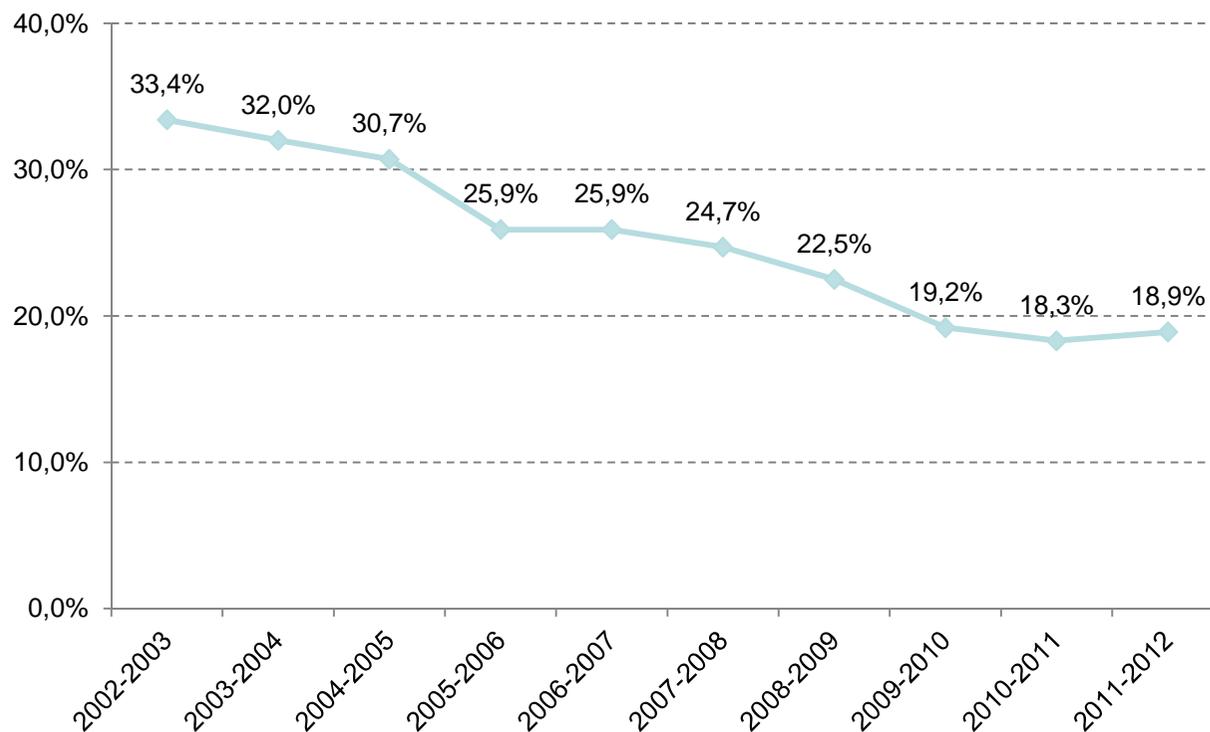
## Taux de numéros restant sans réponse dans l'enquête 126 000 Radio





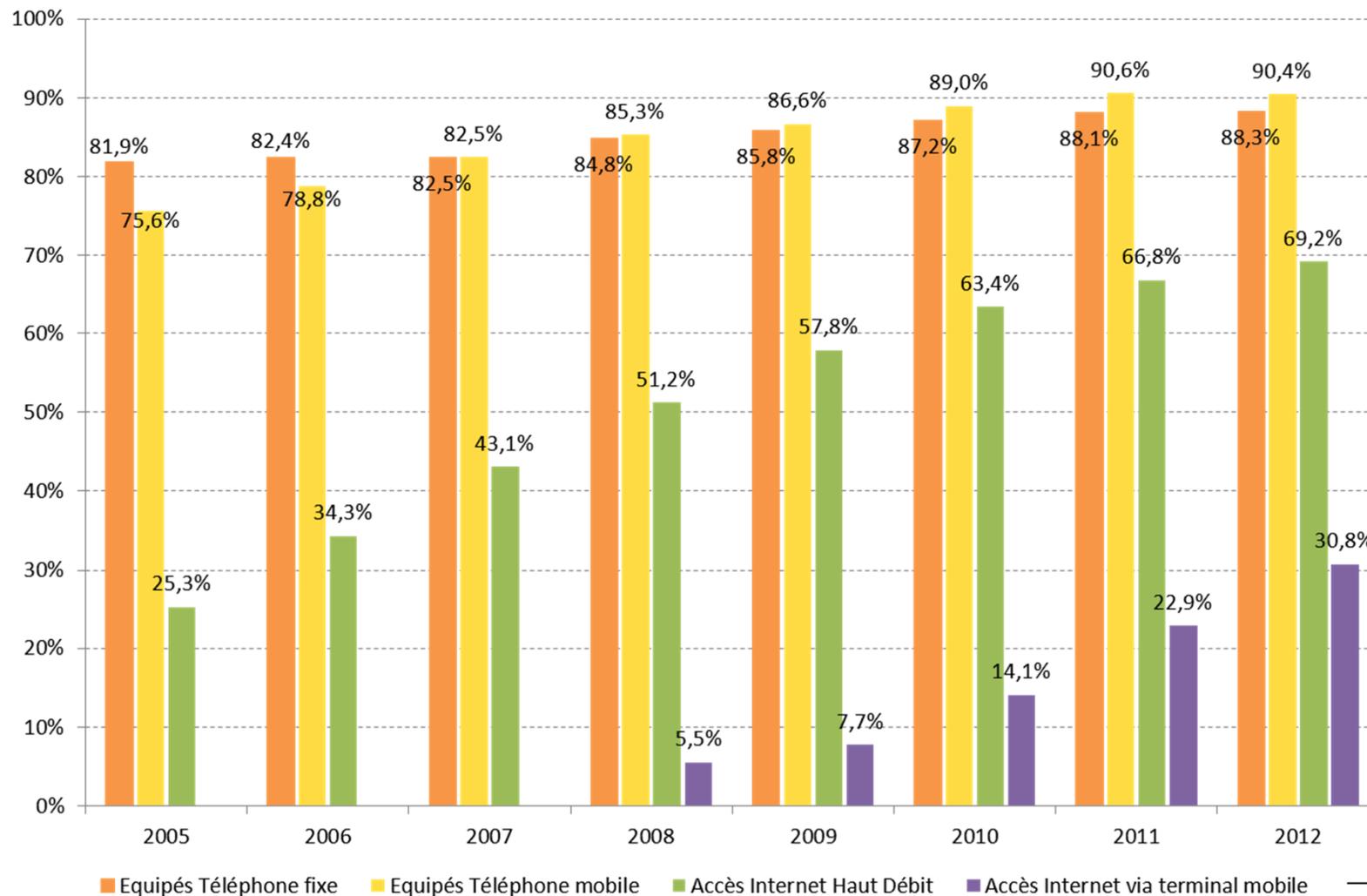
# Baisse de la participation aux panels

## Taux de recrutement pour le Panel Radio





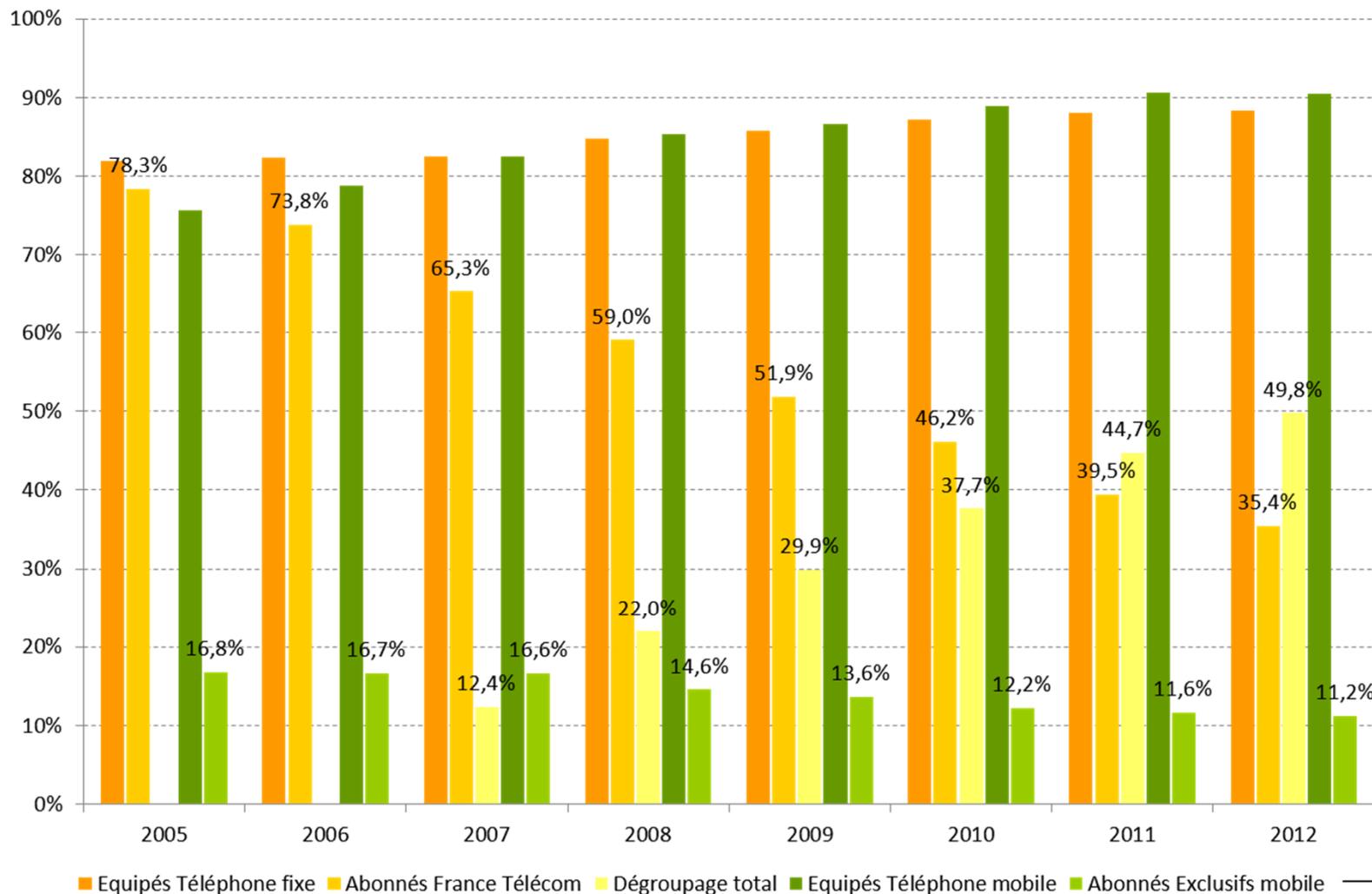
# Evolution des moyens de communication



Source : Médiamétrie – Référence des Equipements Multimédia



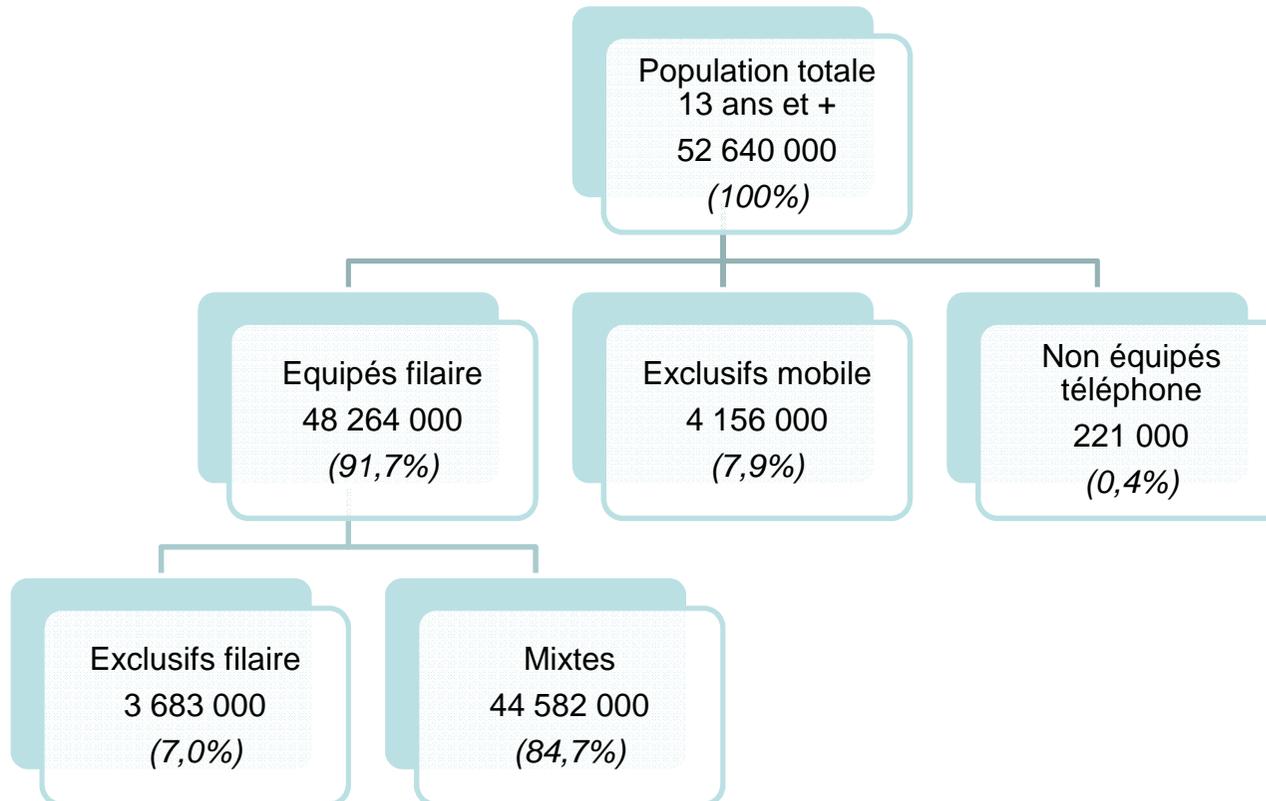
# Focus sur l'équipement téléphonique



Source : Médiamétrie – Référence des Equipements Multimédia



# L'équipement téléphonique aujourd'hui



Source : Médiamétrie – Référence des Equipements Multimédia – Janvier-Juin 2012



# Etude sur la joignabilité comparée entre fixe et mobile

## Méthodologie de l'étude

- ▶ Choix du CAWI pour neutraliser l'effet mode de contact
- ▶ Tirage de 20 000 adresses emails « OK rappel », issus exclusivement d'études réalisées par Médiamétrie
- ▶ Constitution a priori d'un échantillon représentatif des internautes dernier mois sur des critères socio-démographiques, géographique et de fréquence de connexion à Internet
- ▶ Remplissage des questionnaires via Internet pendant 10 jours d'enquête, dont 2 week-ends et 1 relance à mi-parcours

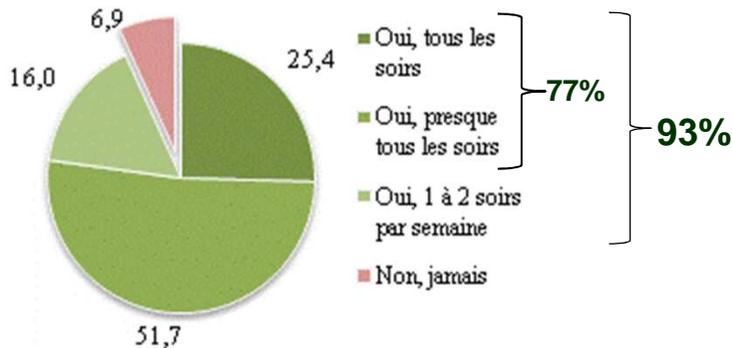
⇒ 1 045 répondants ont validé la totalité du questionnaire online



# Joignabilité comparée fixe vs mobile

« En général, du lundi au vendredi, êtes-vous joignable, vous personnellement, ne serait-ce qu'un moment entre 17h30 et 21h30... ? »

sur un téléphone fixe



sur un téléphone mobile



Base : Double équipés mobile/fixe (962 répondants).

Plus des trois quarts sont joignables tous les soirs ou presque.

93% des double équipés fixe/mobile déclarent être joignables au moins 1 à 2 soirs par semaine sur téléphone fixe (joignabilité globale). Ce taux s'établit à 87% sur téléphone mobile.

Le téléphone fixe reste donc globalement le premier mode de contact de la population, pour réaliser une enquête téléphonique.



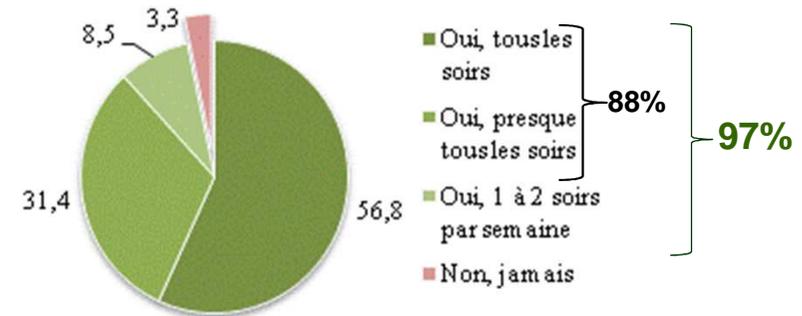
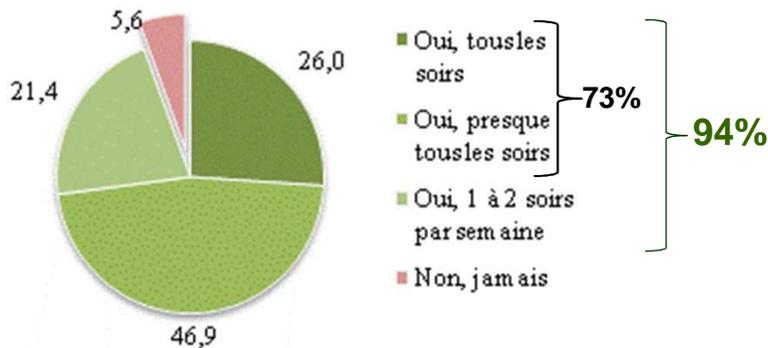
# Les moins de 35 ans, plus joignables sur mobiles

« En général, du lundi au vendredi, êtes-vous joignable, vous personnellement, ne serait-ce qu'un moment entre 17h30 et 21h30... ? »

sur un téléphone fixe

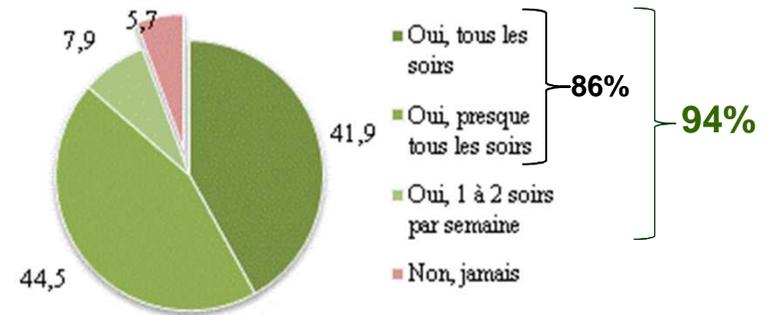
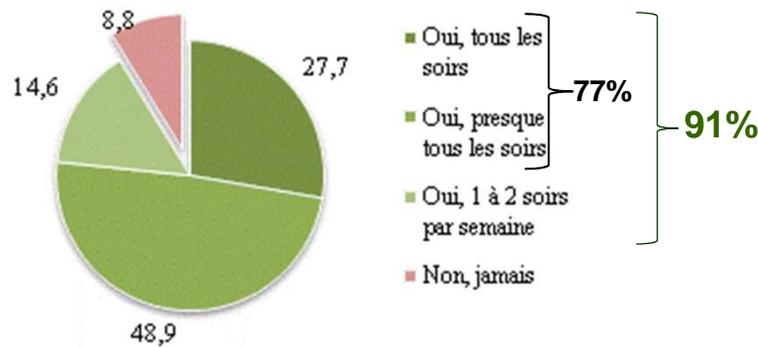
sur un téléphone mobile

15-24 ans



Base : Double équipés mobile/fixe (186 répondants)

25-34 ans

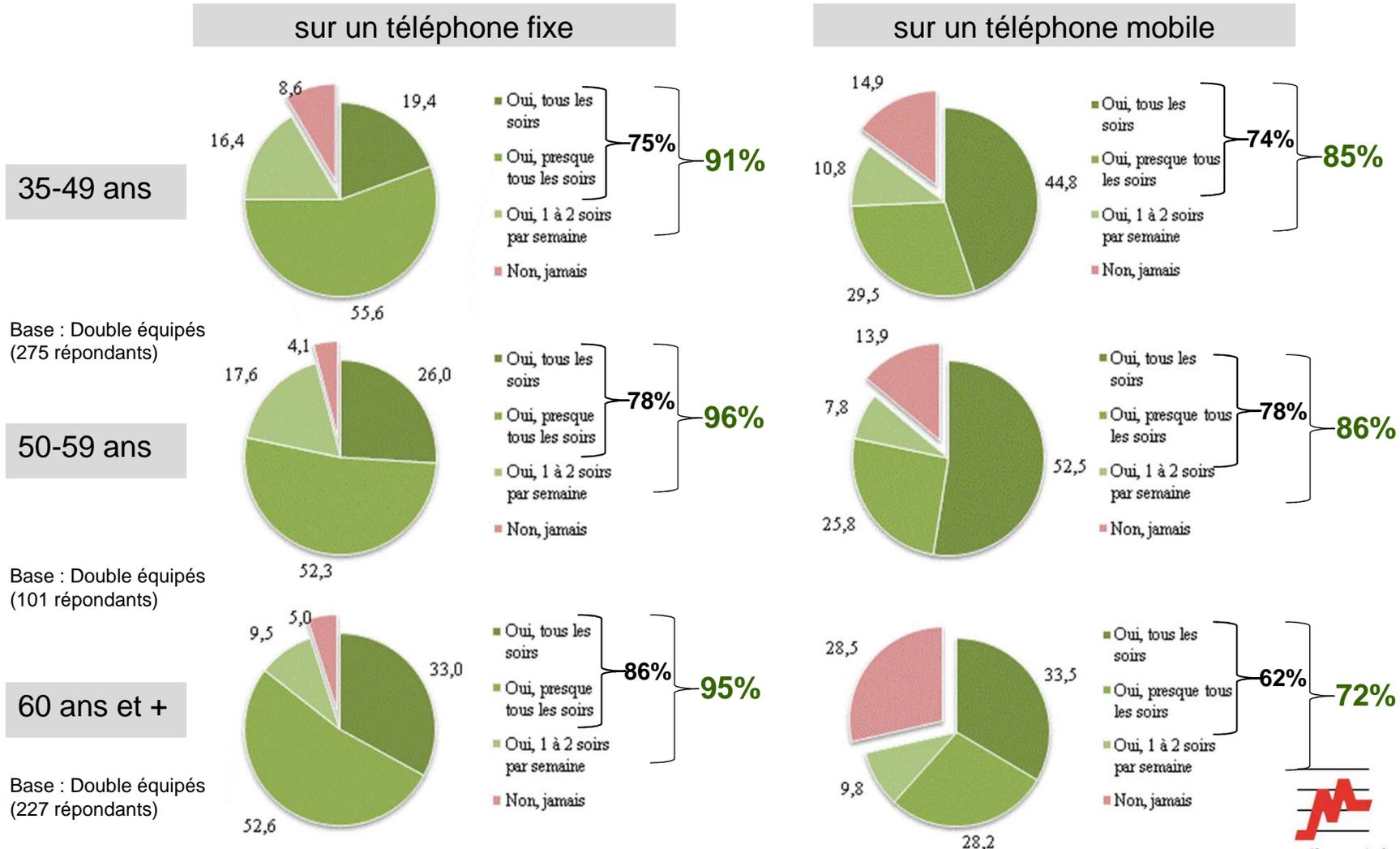


Base : Double équipés mobile/fixe (173 répondants)



# Les plus de 35 ans, globalement plus joignables sur fixe

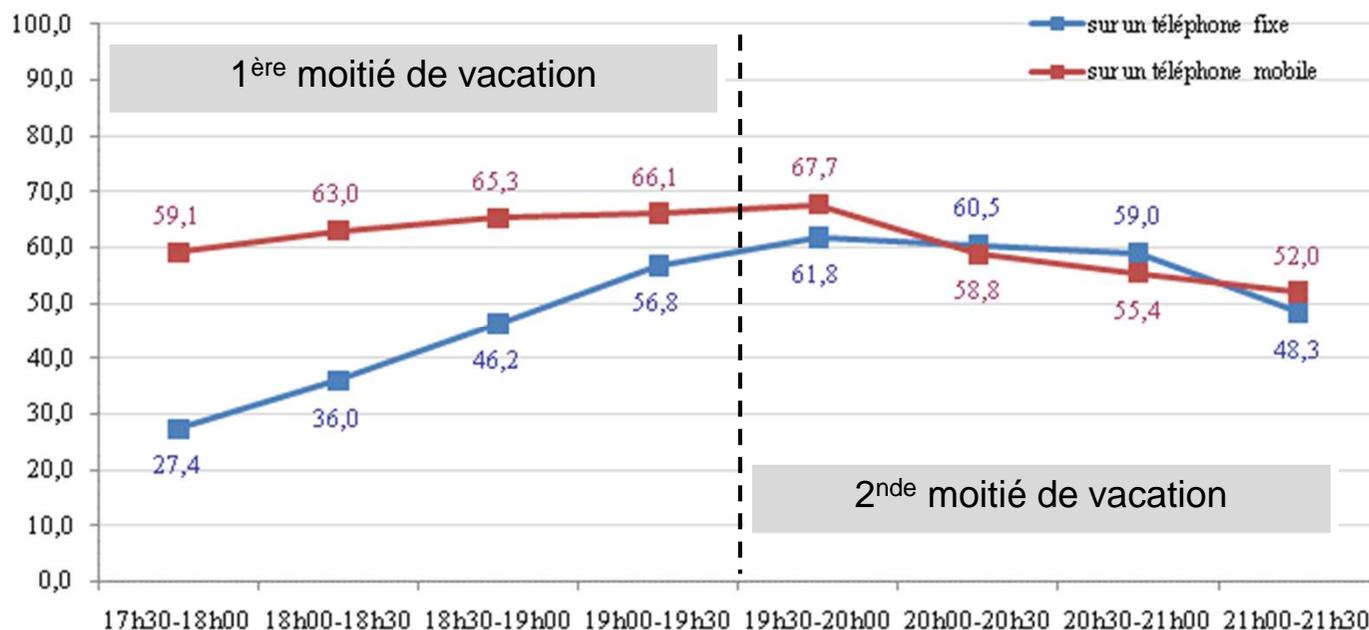
« En général, du lundi au vendredi, êtes-vous joignable, vous personnellement, ne serait-ce qu'un moment entre 17h30 et 21h30... ? »





# Joignabilité comparée fixe vs mobile par tranche horaire

« Et en général, en semaine du lundi au vendredi, à quelle heure êtes-vous joignable, vous personnellement dans votre foyer, sur un téléphone fixe entre 17h30 et 21h30 ? »



Base : Double équipés mobile/fixe (962 répondants)



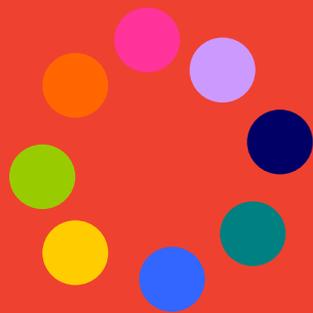
# Grands enseignements sur la joignabilité

## Bilan

- ▶ Le téléphone fixe reste le premier mode de contact et le plus homogène sur toutes les cibles d'âge
  - 93% des équipés fixe et mobile se déclarent joignables en semaine entre 17h30 et 21h30, sur un téléphone fixe
- ▶ Le mobile est un mode de contact très discriminant selon l'âge
  - les raisons de moindre joignabilité sur mobile mettent en évidence la "distance" à l'objet
- ▶ Le mobile permet de joindre encore plus régulièrement certaines populations, notamment les plus jeunes, les CSP+ et en IDF.
- ▶ Toutefois, la joignabilité sur mobile doit être relativisée par la notion de disponibilité

⇒ **Toutes les populations peuvent être interrogées sur téléphone fixe, dans les mêmes proportions, mais le téléphone mobile peut apporter une aide significative pour joindre certaines populations**

# *Les nouveaux protocoles de recueil*





# Pour pallier les erreurs de couverture

## Constitution de la base de sondage dans le cadre des enquêtes téléphoniques

- ▶ Numéros de téléphone fixe géographiques
  - Base de départ : annuaire France Télécom
  - Déclinaison des numéros afin de joindre les numéros en liste rouge
- ▶ Numéros de téléphone mobile (depuis septembre 2003)
  - Objectif : joindre les exclusifs mobile
  - Base de départ : numéros générés aléatoirement
  - Qualification préalable de ces numéros afin d'identifier les individus équipés exclusivement d'un téléphone mobile
- ▶ Numéros de téléphone fixe non géographiques (depuis janvier 2009)
  - Objectif : joindre les équipés filaire en dégroupage total
  - Base de départ : numéros générés aléatoirement
  - Qualification préalable de ces numéros afin d'identifier ceux qui ne sont pas joignables sur un numéro géographique



# Pour pallier les erreurs de non-réponse

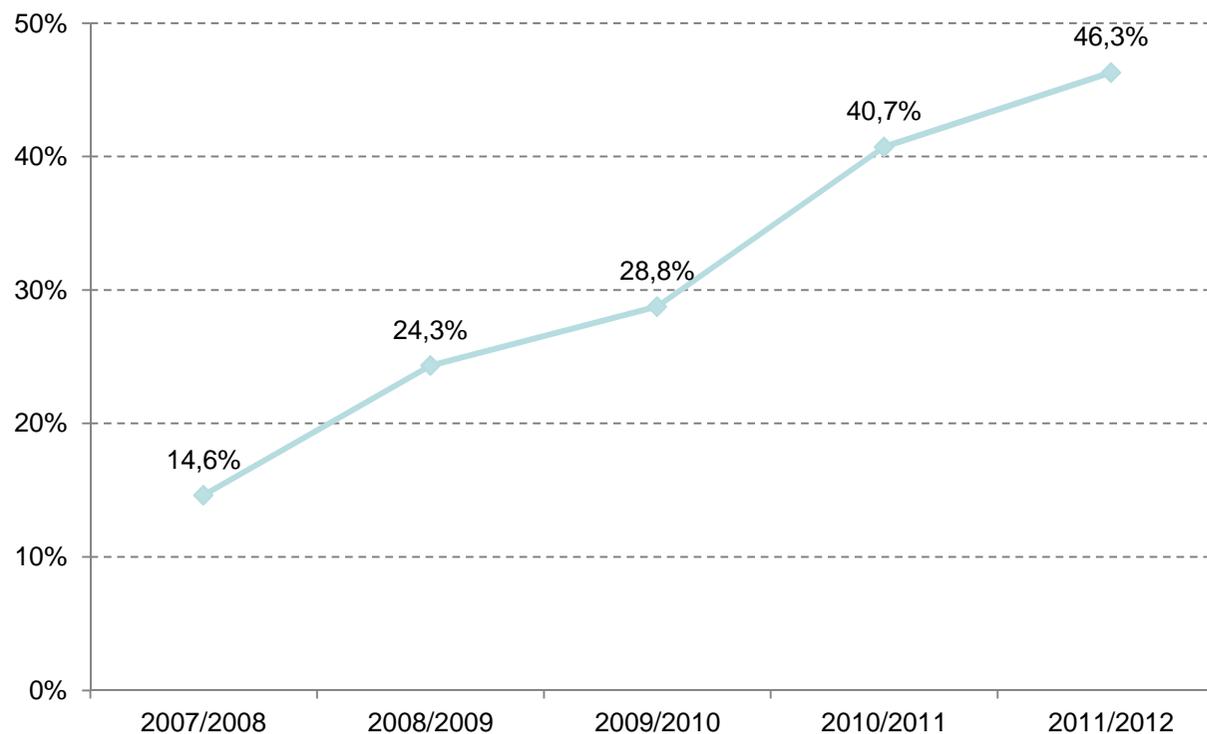
## Cas du Panel Radio : introduction en 2007 du e-Carnet

- ▶ S'adapter à la « zone de confort » des panélistes en leur offrant le choix du mode de remplissage adapté à leurs usages :
  - Internet
  - support papier « traditionnel »
- ▶ Renforcer le taux de maintien des panélistes en cours d'étude, principalement des cibles les plus « abandonnistes », notamment les jeunes hommes
- ▶ Profiter de la démocratisation d'Internet pour développer une nouvelle interface de recueil de l'information



# Pour pallier les erreurs de non-réponse

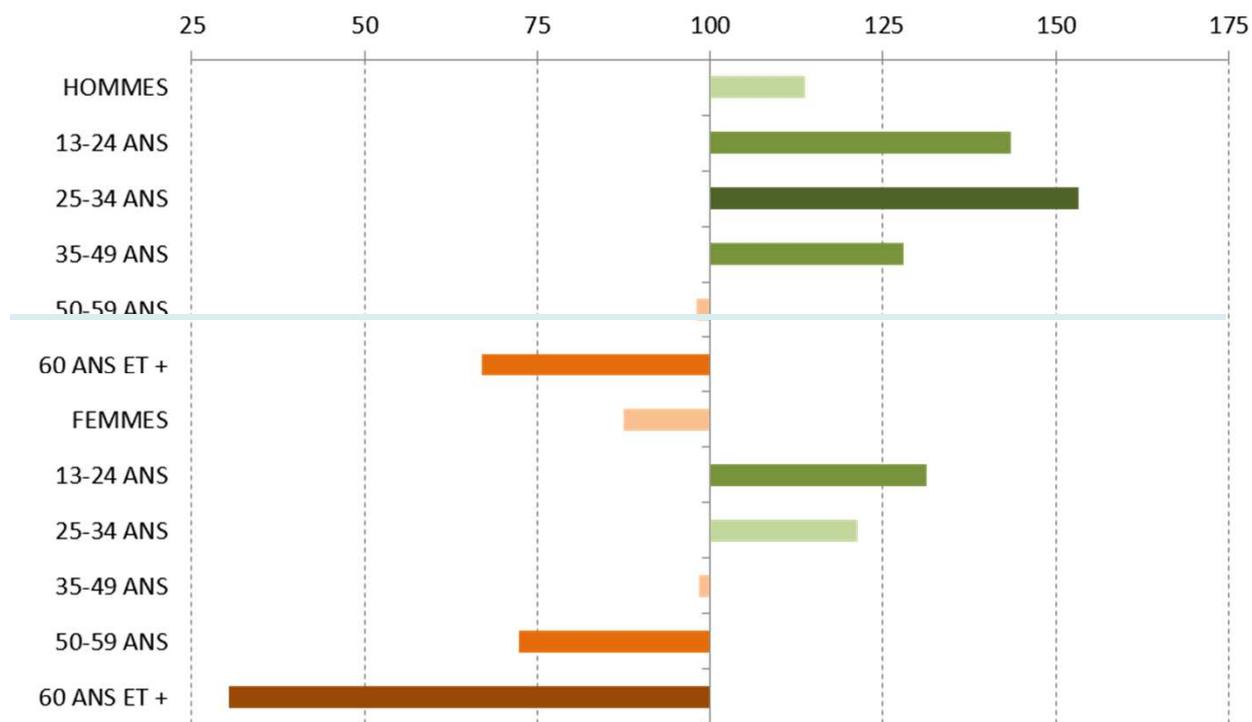
## Evolution de la part de e-Panélistes





# Pour pallier les erreurs de non-réponse

## Profil des e-Panélistes comparé à l'ensemble des panélistes





# Pour pallier les erreurs de non-réponse

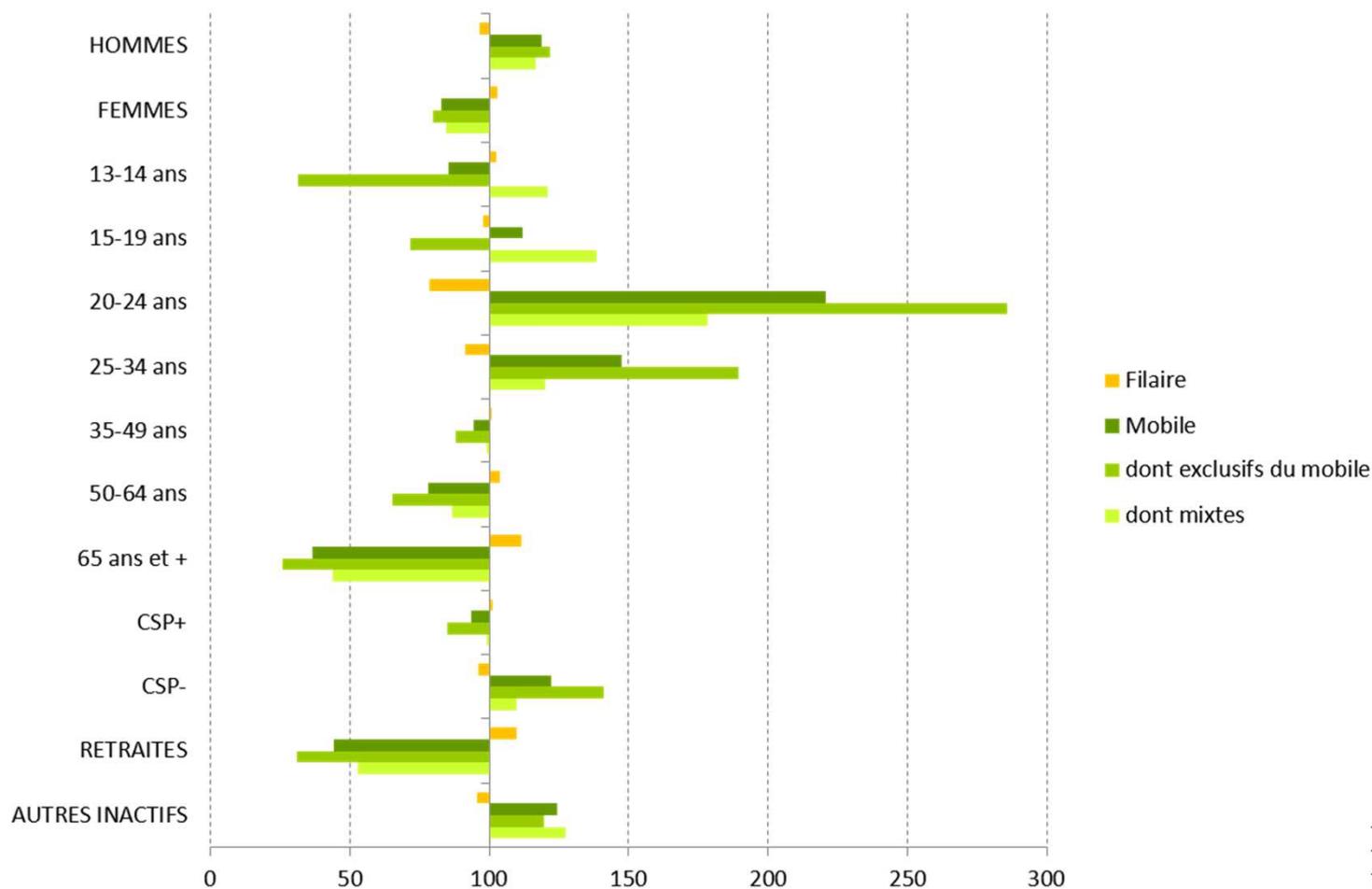
## Cas de la 126 000 Radio

- ▶ Introduction en janvier 2012 d'interviews par téléphone mobile d'individus non exclusifs du mobile
- ▶ Effet direct sur la qualité du recueil :
  - De 14 à 11 numéros nécessaires pour réaliser une interview
  - Des vacations raccourcies car des quotas plus faciles à respecter

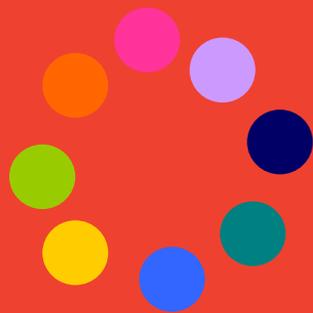


# Pour pallier les erreurs de non-réponse

## Profil des répondants par type d'appel (en indice)



# *Les conséquences sur l'échantillon*





# Le problème des probabilités d'inclusion

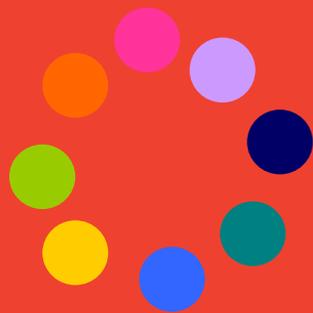
## La question soulevée

- ▶ La dernière évolution mise en œuvre dans le cadre de l'enquête 126 000 Radio confère à un individu multi-équipé fixe et mobile une probabilité plus forte d'être interrogé dans l'enquête
- ▶ Comment neutraliser cet effet ?

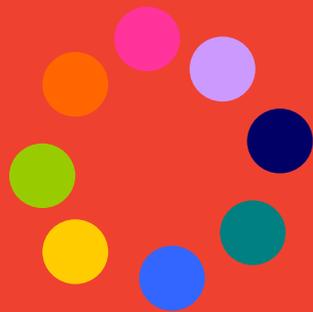
## Les éléments de réponse

- ▶ Négligeable du fait des taux de sondage très faibles
- ▶ Redressement de l'échantillon sur un critère d'équipement téléphonique

# Les mesures hybrides



# *Définition*

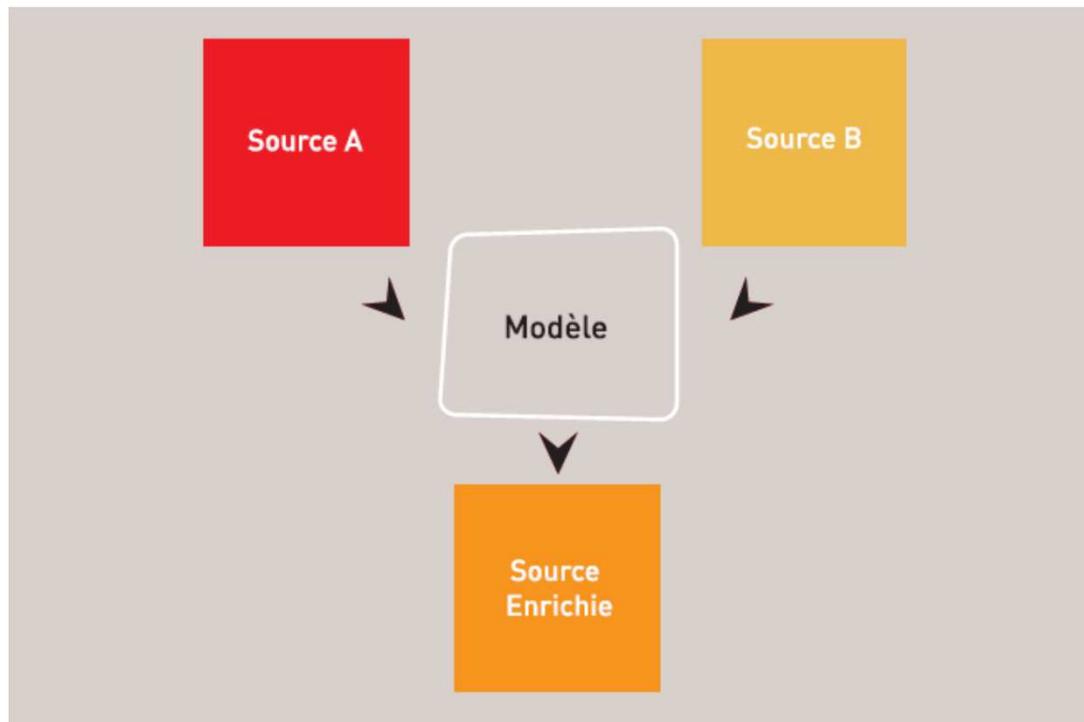




# Qu'est-ce que la mesure hybride ?

## Mixer "sondages" et "exhaustivité"

- ▶ Mélanger deux sources d'information de natures et de niveaux différents, les croiser mutuellement pour en créer une troisième, plus fine ou plus riche





# Qu'est-ce que la mesure hybride ?

## Origine des applications à la mesure d'audience

- ▶ Le développement des technologies numériques et des nouveaux moyens de communication offrent de nouvelles perspectives à la mesure d'audience
  - multiplication des bases de données exhaustives telles les données des Box en TV, des téléphones mobiles ou de fréquentation des sites Internet
  
- ▶ Même si ces informations sont souvent de granularité différente ou couvrent un périmètre partiel, il est naturel de vouloir les utiliser afin d'améliorer la précision des résultats



## Des informations auxiliaires de plus en plus nombreuses

### **Disposer d'information auxiliaire est a priori un atout**

- ▶ Utiliser une information auxiliaire permet d'améliorer la précision des estimateurs
- ▶ Cette information auxiliaire peut être utilisée au niveau de l'échantillonnage ou au niveau du calcul des estimateurs

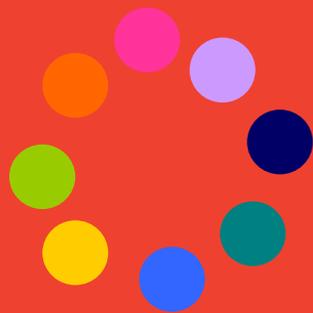
### **Jusqu'à présent, les seules informations auxiliaires dont on disposait pour nos enquêtes provenaient de l'INSEE**

- ▶ Elles étaient utilisées en amont (nb d'interviews par commune, quotas) et en aval (redressement)

### **Comment utiliser ces informations complémentaires ?**

- ▶ La question n'est pas tant celle de l'amélioration de la précision des enquêtes que celle de la mise en cohérence des résultats de sources différentes

# *Méthodes proposées*





## Deux grands types d'approche

### L'approche « panel-up »

- ▶ La donnée de base est celle de l'enquête
- ▶ La donnée exhaustive est utilisée en tant que variable auxiliaire pour améliorer la précision des résultats de l'enquête

### L'approche « log-up »

- ▶ La donnée exhaustive est le socle d'analyse
- ▶ Les résultats de l'enquête sont utilisés pour enrichir la mesure exhaustive (profil,...)



## L'approche panel-up

### **La méthode classique : le calage sur marges**

- ▶ Le principe consiste à modifier les poids de sondage de manière à faire coïncider les totaux observés sur l'échantillon avec ceux calculés sur la population
- ▶ L'utilisation de la macro CALMAR, mise au point par O. Sautory (INSEE) sur l'approche générale formulée par J.C. Deville et C.E. Särndal (1992), permet de minimiser la distance entre les poids de redressement et les poids de sondage



## L'approche panel-up

### **Le problème de la multiplicité des variables auxiliaires**

- ▶ Combien de variables auxiliaires peut-on faire intervenir ?
- ▶ Au-delà de la question de la convergence de l'algorithme se pose celle de la dispersion des poids de redressement
- ▶ Dans le cas d'une trop grande dispersion des poids de redressement, l'estimateur redressé peut perdre son efficacité

**Trouver un compromis acceptable !**



# L'approche log-up : le cas de l'Internet mobile

## Deux sources d'information

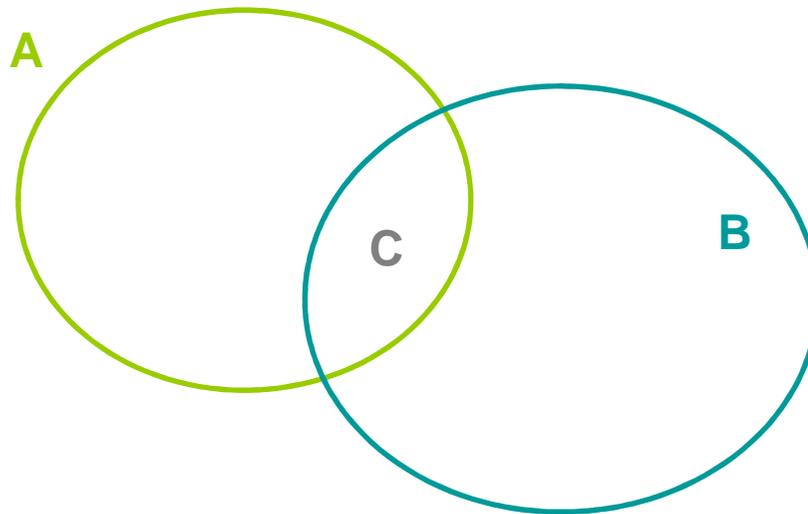
- ▶ Les logs des opérateurs de téléphonie
  - Nombre d'individus connectés par site/application
  - Nombre de pages vues par site
- ▶ Un panel de 10 000 individus
  - Profil des visiteurs des sites/applications
  - Duplication entre sites/applications

## Comment combiner ces deux sources ?



# L'approche log-up : le cas de l'Internet mobile

## L'estimation de la duplication



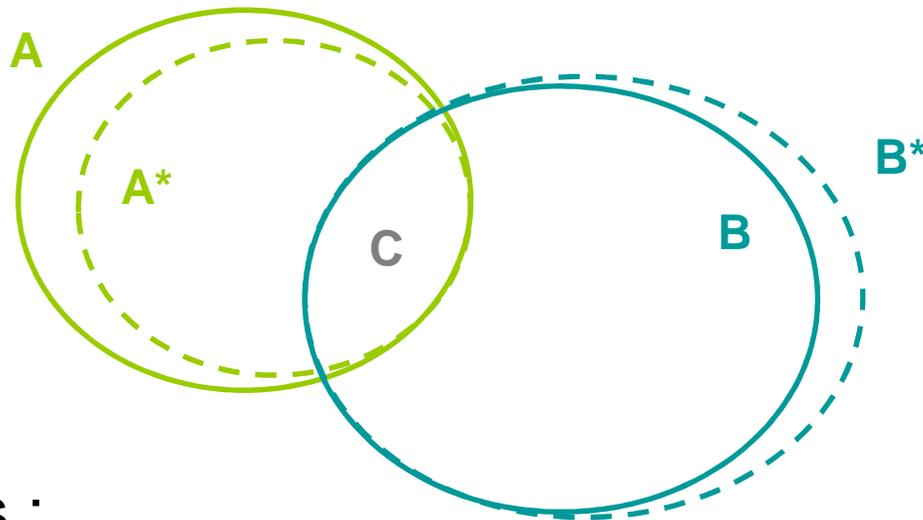
### ► Notations :

- A, B le nombre de visiteurs respectifs des sites A et B dans le panel
- C le nombre de visiteurs communs aux sites A et B dans le panel



# L'approche log-up : le cas de l'Internet mobile

## L'estimation de la duplication



► Notations :

- $A^*$ ,  $B^*$  le nombre de visiteurs respectifs des sites A et B dans la mesure exhaustive

► Question : estimation de  $C^*$



# L'approche log-up : le cas de l'Internet mobile

## Les hypothèses de départ

- ▶ Dans le cas où A et B sont indépendants, c'est-à-dire

$$C = \frac{A \times B}{N}, \text{ alors on veut que } C^* = \frac{A^* \times B^*}{N} = C_{\text{ind}}$$

- ▶ Dans le cas où A est inclus dans B, alors

$$C^* = \min\left(\frac{C \times B^*}{B}, A^*\right) = C_{\text{inc}}$$

- ▶ Dans les cas intermédiaires, on est entre les deux...



# L'approche log-up : le cas de l'Internet mobile

## Notations

▶  $K = \min\left(\frac{C}{A \times B/N}, \frac{A \times B/N}{C}\right)$

▶  $L = \frac{C}{A}$

▶ Alors  $C^*$  est estimée par : 
$$C^* = \frac{C_{\text{ind}} \times (1-L) + C_{\text{inc}} \times (1-K)}{1-L+1-K}$$



# L'approche log-up : le cas de la voie de retour TV

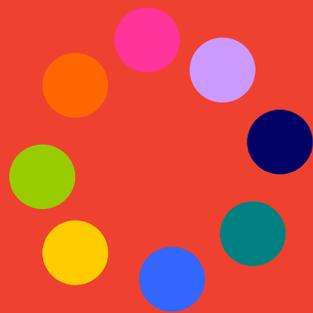
## Deux sources d'information

- ▶ Les logs des opérateurs TV avec voie de retour
  - Nombre de décodeurs allumés et la chaîne
- ▶ Un panel de 5000 foyers
  - Nombre de postes allumés et chaîne regardée quelque soit la source
  - Individus présents devant l'écran

## Comment combiner ces deux sources ?

- ▶ Etape 1 = Décodeur → TV
  - Ecrêtage des audiences
- ▶ Etape 2 = TV → Individu
  - Individualisation des données

# Perspectives et enjeux en matière de R&D





# Les perspectives

## Travaux en cours sur les méthodes de calage

- ▶ Calage sur composantes principales
- ▶ Calage pénalisé et méthode ridge

## Traitement des big data

- ▶ Les mesures exhaustives sont difficiles à exploiter de par leur volumétrie
  - Cloud computing
  - Hadoop



# La statistique d'enquête aujourd'hui et demain

## Constats

- ▶ L'observation seule ne permet plus de répondre aux problématiques actuelles de la mesure d'audience
- ▶ On s'éloigne progressivement du cadre « classique » de la statistique d'enquête pour aller vers des systèmes « hybrides » mixant observation et modélisation

## Risques

- ▶ Ne pas négliger pour autant la qualité de l'échantillonnage
- ▶ Le recours à la modélisation ne permet pas de corriger les défauts de l'observation

**Merci pour votre attention**

**[avanheuverzwyn@mediametrie.fr](mailto:avanheuverzwyn@mediametrie.fr)**

