

COMPROMIS ENTRE ERREUR DE NON-RÉPONSE ET ERREUR DE MESURE DANS COSET-MSA

Gaëlle Santin^{1,2}, Pauline Delézire¹, Béatrice Geoffroy¹, Laetitia Bénézet¹, Jean Bouyer³, Alice Guéguen²

¹Institut de veille sanitaire – Département santé travail, 12 rue du Val d’Osne 94415 Saint-Maurice Cedex

²UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population", 16 avenue Paul Vaillant Couturier, 94807 Villejuif Cedex

³CESP-INSERM 1018 - Equipe 2, 82 rue du Général Leclerc, 94276 Le Kremlin-Bicêtre Cedex

La surveillance épidémiologique

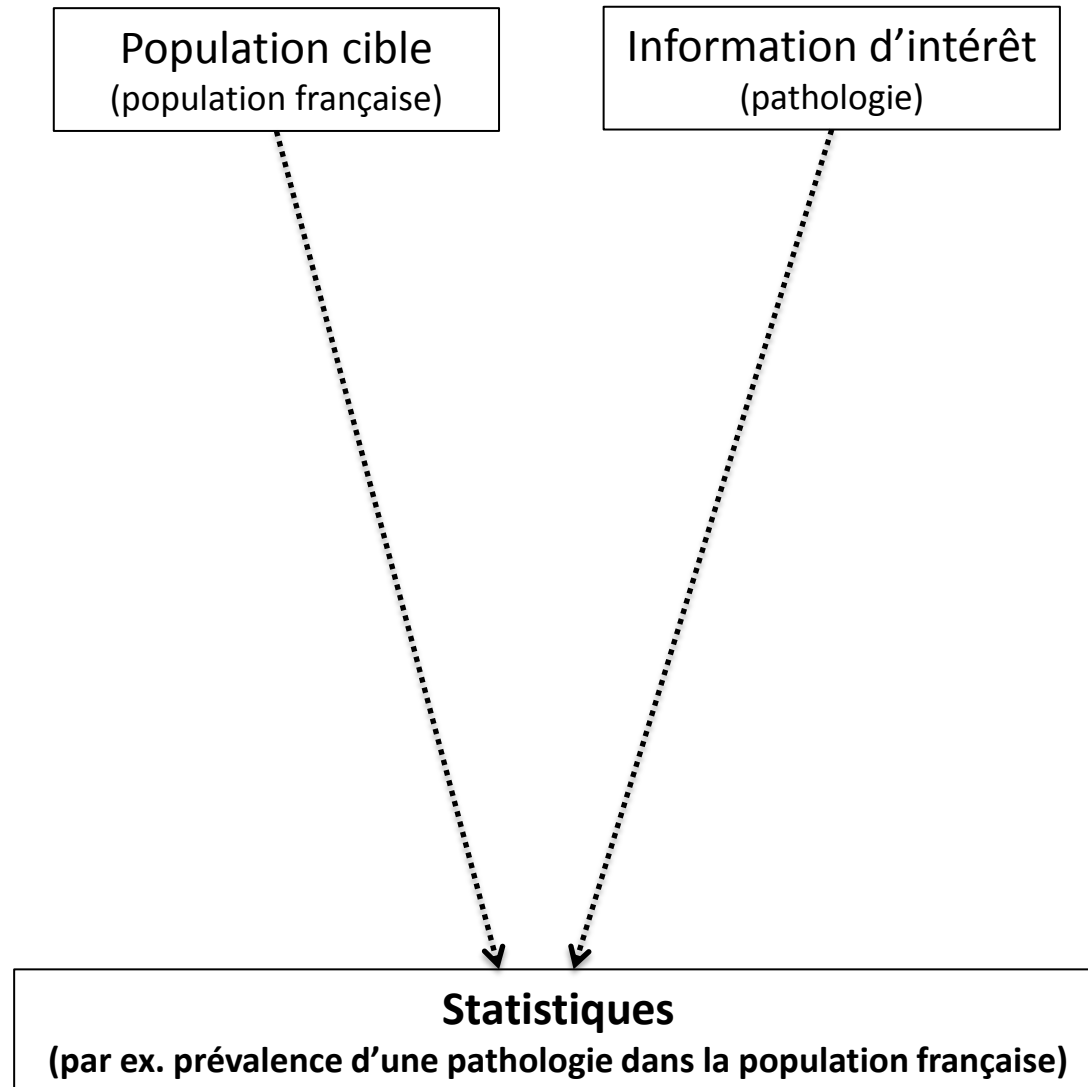
- **Définition**

Suivi et analyse épidémiologique systématiques et permanents d'un problème de santé et de ses déterminants à l'échelle d'une population, afin de les contrôler par des interventions au niveau collectif, et d'identifier des phénomènes inconnus en termes d'effets ou de déterminants (*Goldberg, Imbernon 2008*)

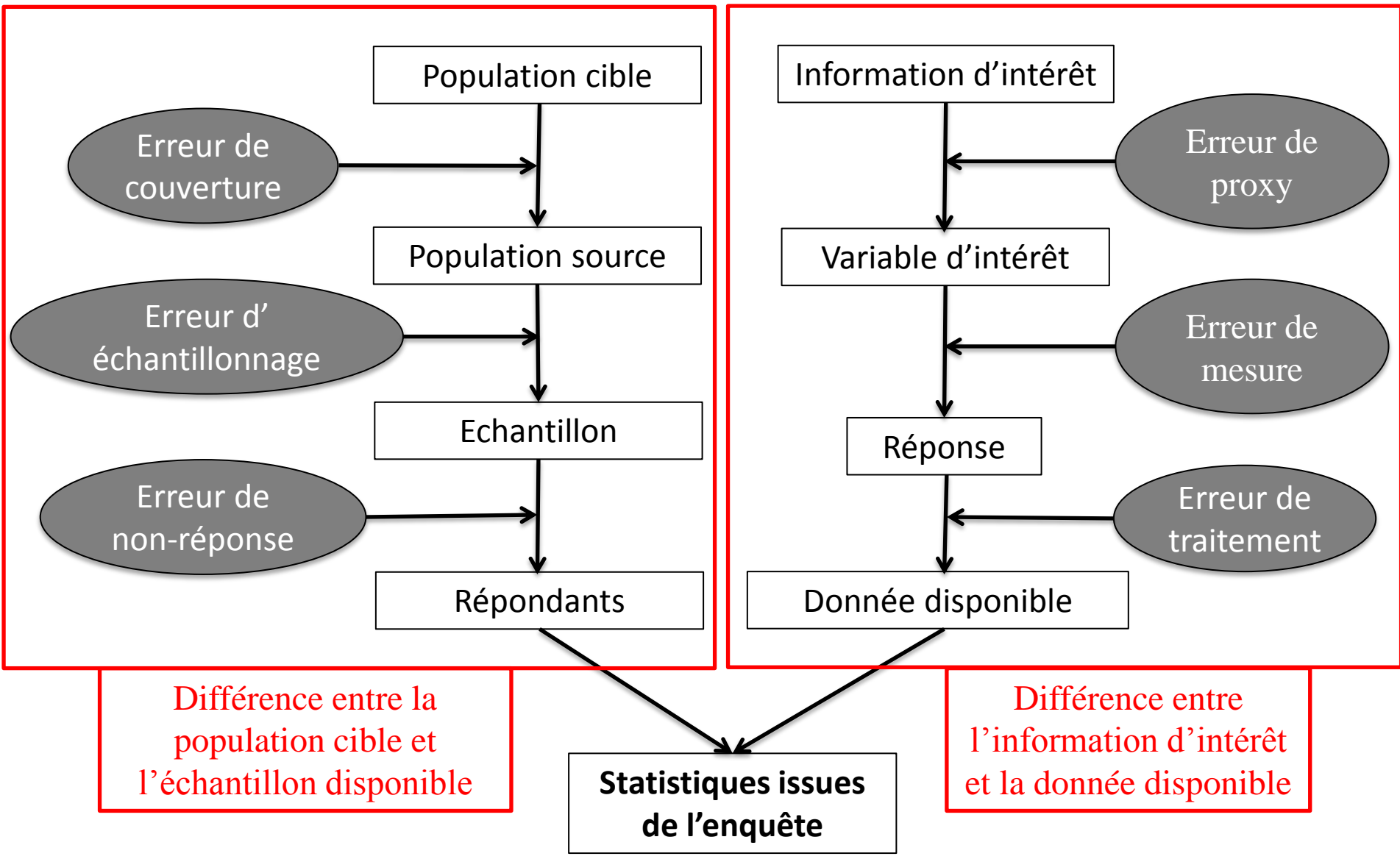
- **Systèmes de surveillance**

- Remontées d'information issues de systèmes existants
- Informations collectées avec des objectifs épidémiologiques spécifiques : registres ou **enquêtes en population**

L'erreur totale dans les enquêtes



L'erreur totale dans les enquêtes



L'erreur totale dans les enquêtes

- Soit Y la variable d'intérêt
- Soit θ le paramètre d'intérêt (moyenne, prévalence) et $\hat{\theta}$ son estimateur
- Erreur totale : $ETot(\hat{\theta}) = \hat{\theta} - \theta$

↳ Mesurée par l'Erreur Quadratique Moyenne

$$EQM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \quad \longleftrightarrow \quad EQM(\hat{\theta}) = \text{Biais}^2(\hat{\theta}) + \text{Var}(\hat{\theta})$$

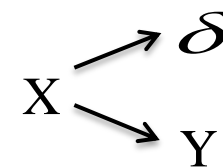
	Variance faible	Variance élevée
Absence de biais		
Présence de biais		

▪ Estimation du paramètre sur un échantillon

● Vraie valeur du paramètre

Biais de non-réponse : comment les minimiser ?

- **Augmenter le taux de réponse aux enquêtes**
 - Taux de réponse global
 - Taux de réponse par groupe d'individus
 - **Enquête en deux phases pour non-réponse**
- **Déterminer un estimateur qui prend en compte les causes communes de la probabilité de réponse et de la variable d'intérêt**
 - Identifier les causes communes X liées à la fois à la probabilité de réponse δ et à la variable d'intérêt Y
 - Recueillir les causes communes X chez les répondants et les non-répondants
 - Imputation
 - **Repondération**



Biais de mesure : comment les minimiser ?

- Adapter le mode de collecte des données à la thématique de l'enquête ou à la personne à enquêter :
par enquêteur ou questionnaire auto-administré
- Limiter la longueur et la complexité du questionnaire
- Réfléchir à l'ordre des questions :
effet de halo ou de contamination

Taux de réponse, biais de non-réponse, biais de mesure

- **Hypothèses relatives à un taux de réponse élevé :**
 - Diminution du biais de non-réponse : mais hypothèse discutée (*Groves 2006*)
 - Augmentation du biais de mesure : réponse moins précise ou volontairement erronée des personnes auprès de qui on fait beaucoup d'efforts pour obtenir une réponse (*Stang 2004*)
 - Utilité d'atteindre un taux de réponse maximal ?
- **Peu d'études ayant étudié quantitativement ce problème** (*Peytchev 2010 ; Olson 2006 Kreuter 2012*) :
 - Erreur=biais
 - Quantification de l'EM et de l'ENR selon la propension à répondre ou la difficulté à joindre les personnes
 - Résultats différents selon les variables étudiées
 - Sans correction de la non-réponse

Objectif

Etudier le compromis entre erreur de non-réponse et erreur de mesure selon la difficulté à joindre les personnes dans Coset-MSA :

- sans correction de la non-réponse
- avec correction de la non-réponse

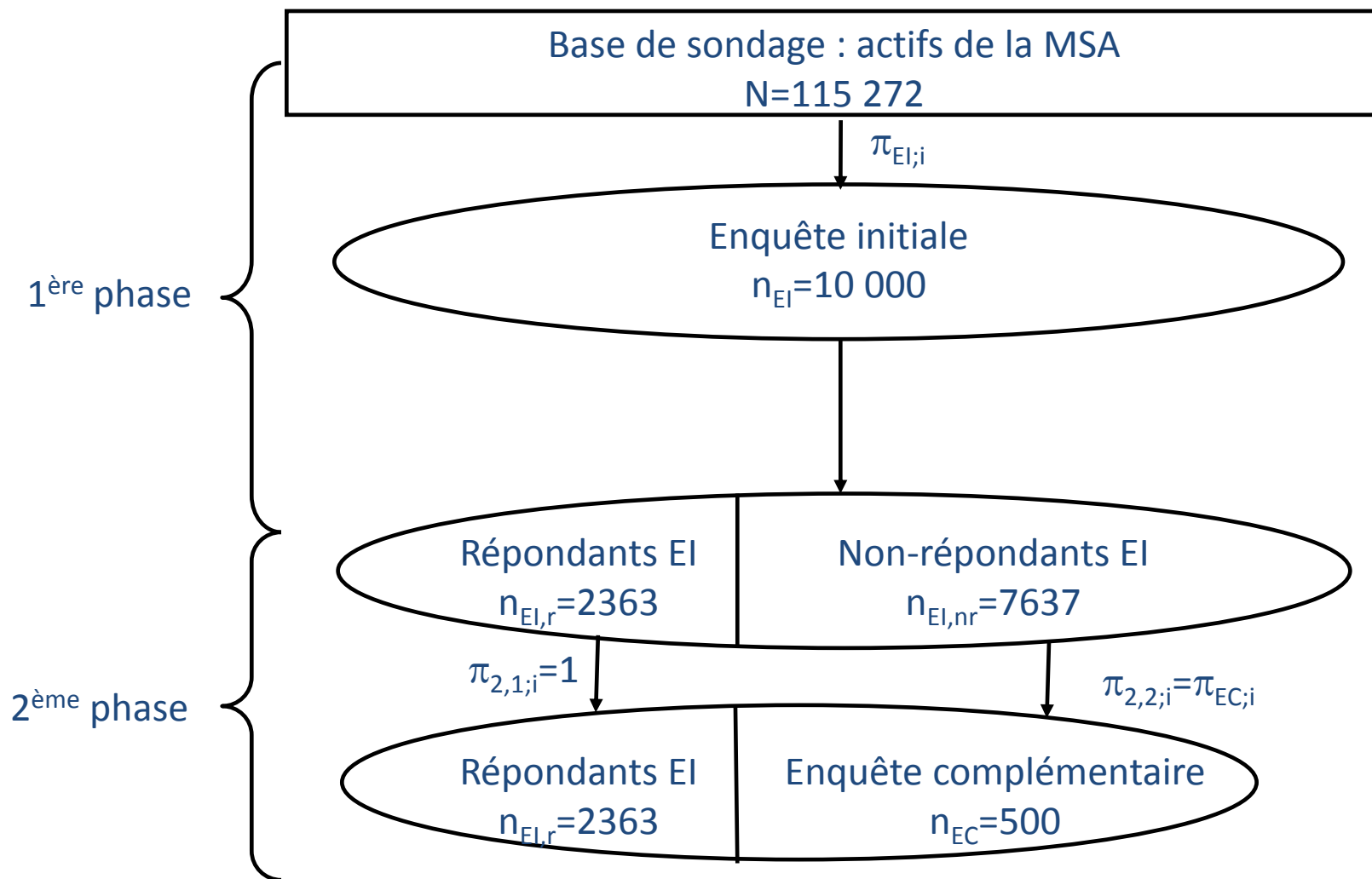
Echantillonnage

Phase pilote Coset-MSA – Inclusion
Cohorte pour la Surveillance Épidémiologique en lien avec le Travail
Chez les travailleurs affiliés à la Mutualité Sociale Agricole (MSA)

- **Population éligible**
 - Salariés et non salariés agricoles
 - 18 à 65 ans
 - Ayant travaillé au moins 90 jours en 2008
 - 5 départements pilotes (Bouches-du-Rhône, Pas de Calais, Pyrénées Atlantique, Saône et Loire, Finistère)
- **Base de sondage** : Base d'assurance retraite de la MSA

Echantillonnage

Enquête en deux phases pour non-réponse (*Hansen-Hurwitz 1946*)
(ou enquête en deux phases)



Données recueillies (1)

	Enquête initiale (EI)	Enquête complémentaire (EC)
Période d'enquête	Février-mai 2010	Novembre 2010-février 2011
Mode de recueil des données	Autoquestionnaire postal	Téléphone ou face-à-face
Données recueillies par questionnaire		
. Répondants	Santé perçue, échelles de santé (TMS, symptômes dépressifs, asthme), consommation de tabac et d'alcool, historique d'emplois et d'expositions professionnelles	[Questionnaire court issu de l'EI] Santé perçue, consommation de tabac et d'alcool, emploi actuel, quelques expositions professionnelles actuelles
. Taux de réponse	24%	63%
	70%	

Données recueillies (2)

Données disponibles via les systèmes d'information (SI) existants : **Informations auxiliaires**

. Répondants et non-répondants

Variables sociodémographiques (2008)

Données de santé : Assurance maladie, SNIIR-AM (2008 à 2010)

- *Remboursement de soins*
- *Absentéisme pour raison de santé - Indemnités journalières*
- *Hospitalisations*

Données professionnelles : Cotisations professionnelles, MSA

- *Dernier emploi principal en date (2008, 2009 ou 2010)*
- *Accidents du Travail et Maladies Professionnelles (ATMP) reconnus (2003 à 2008)*

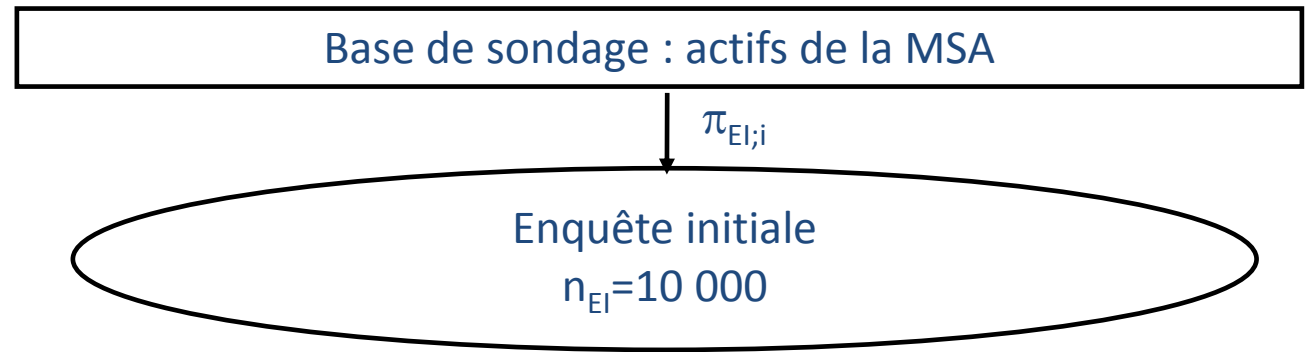
Estimation des prévalences

$$\hat{p} = \frac{\sum w_i y_i}{\sum w_i}$$

Echantillon complet
Enquête initiale
Enquête en 2 phases

Estimation des prévalences

$$\hat{p} = \frac{\sum w_i y_i}{\sum w_i}$$



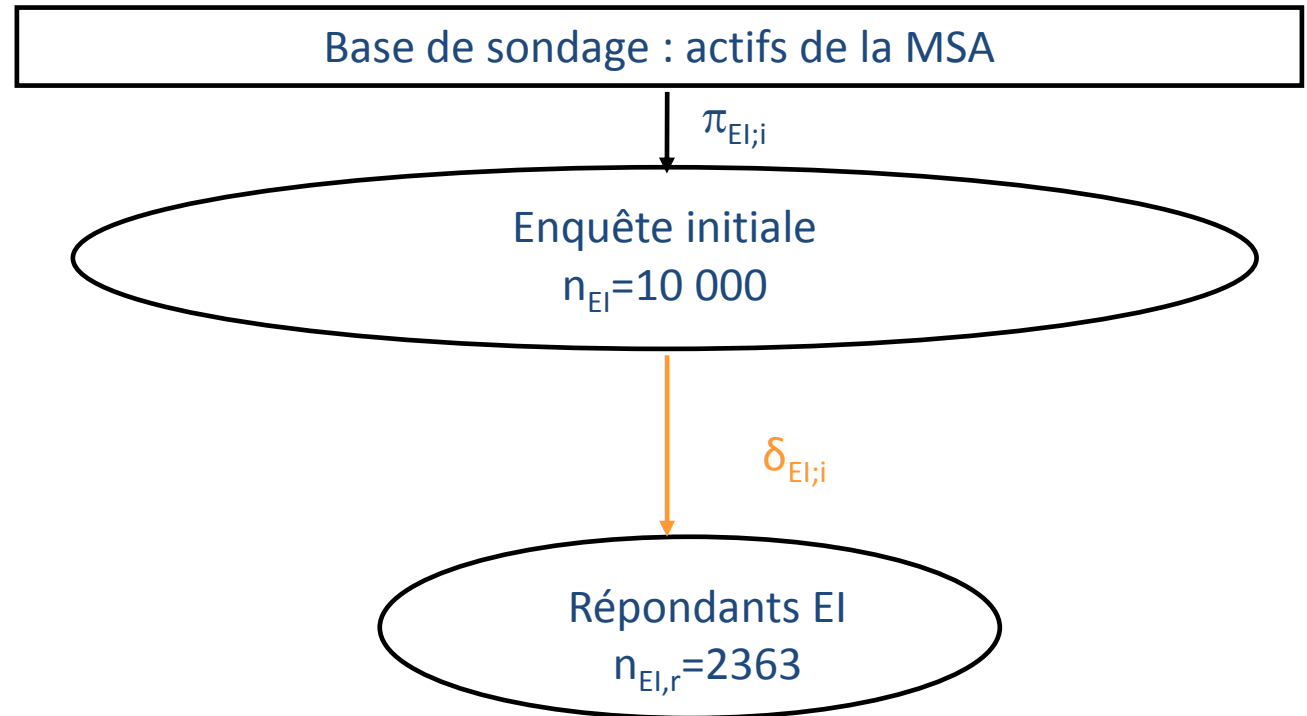
Echantillon complet

Enquête initiale

Enquête en 2 phases

Estimation des prévalences

$$\hat{p} = \frac{\sum w_i y_i}{\sum w_i}$$

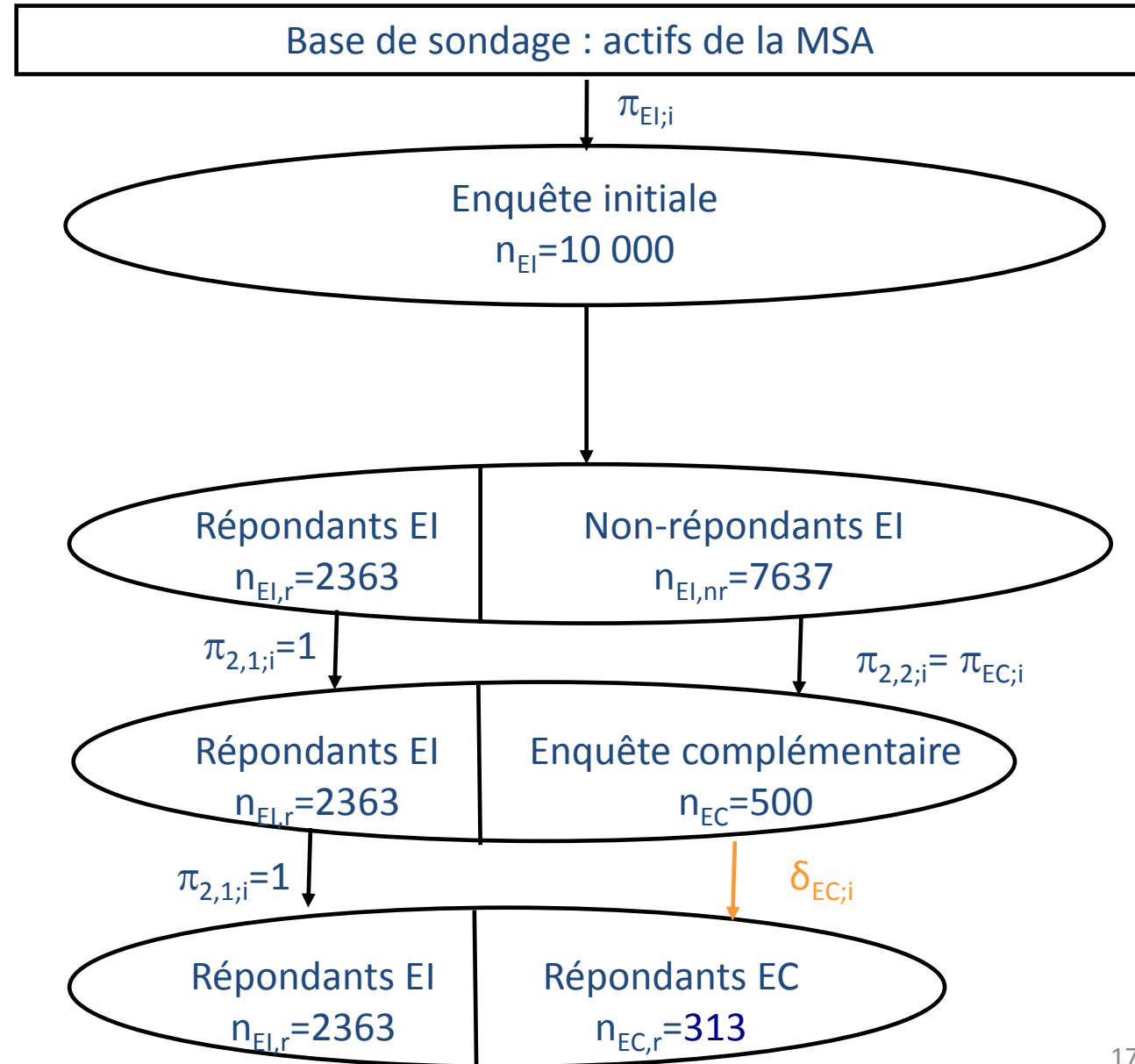


Echantillon complet
Enquête initiale
 Enquête en 2 phases

Estimation des prévalences

$$\hat{p} = \frac{\sum w_i y_i}{\sum w_i}$$

Echantillon complet
 Enquête initiale
Enquête en 2 phases



Correction de la non-réponse

- **Probabilités de réponse δ_{EI} et δ_{EC} inconnues**
 - Les estimer ($\hat{\delta}_{EI}$ et $\hat{\delta}_{EC}$) sous certaines hypothèses
- **Hypothèse MCAR : taux de réponse moyen**
 - $\hat{\delta}_{EI} = 24\%$ et $\hat{\delta}_{EC} = 63\%$
- **Hypothèse MAR conditionnellement aux informations auxiliaires X :**
 - X : variables sociodémographiques, du SNIIRAM et de la MSA
 - ++ des informations auxiliaires : directement liées à la thématique de l'enquête (santé et travail)**
 - Méthode des scores (*Little 1986 ; Eltinge, Yansaneh 1997 ; Haziza, Beaumont 2007*)

Principe général de la méthode (Olson 2006)

- **Idée** : Quantifier les erreurs de mesure et les erreurs de non-réponse selon la difficulté à joindre les personnes
- **Données nécessaires** :
 - Variable gold standard comprenant mesurée pour l'ensemble des personnes tirées au sort (issue de systèmes d'information existants, notée « si »)
 - Même variable mesurée par questionnaire, donc uniquement chez les répondants (notée « qaire »)

		Variable d'intérêt Y	
		Système d'information Y_{si}	Questionnaire Y_{qaire}
Echantillon	1	$Y_{si,1}$	$Y_{qaire,1}$
	·		
	·		
	n_r	Y_{si,n_r}	Y_{qaire,n_r}
	·		
	·		
	n	$Y_{si,n}$	

Répondants

Principe général de la méthode

- **Erreur de non-réponse (ENR) :**

$$\hat{E}_{NR} = \hat{y}_{s,si} - \hat{y}_{sr,si}$$

- **Erreur de mesure (EM) :**

$$\hat{E}_M = \hat{y}_{sr,si} - \hat{y}_{sr,qaire}$$

- **Erreur totale (ET) :**

$$\hat{E}_{tot} = \hat{E}_{NR} + \hat{E}_M$$

ENR, EM, ET estimées :

- Pour les personnes faciles à joindre
- Puis en rajoutant les personnes moins faciles à joindre
- Jusqu'à inclusion de l'échantillon complet de répondants

	Système d'information y_{si}	Système d'information y_{si}	Questionnaire y_{qaire}
1	$\hat{y}_{s,si}$	$\hat{y}_{sr,si}$	$\hat{y}_{sr,qaire}$
.			
.			
n_r			
.			
.			
n			

Application au pilote Coset-MSA

- **Quelle définition de la difficulté à joindre les personnes ?**
 - Faible = répondant à l'enquête initiale (EI)
 - Elevée = répondant à l'enquête complémentaire (EC)
- **Quelles données utiliser ?**

Disposer de variables identiques dans les données gold standard et dans les données de questionnaire

→ Seulement dans les SI de la MSA :

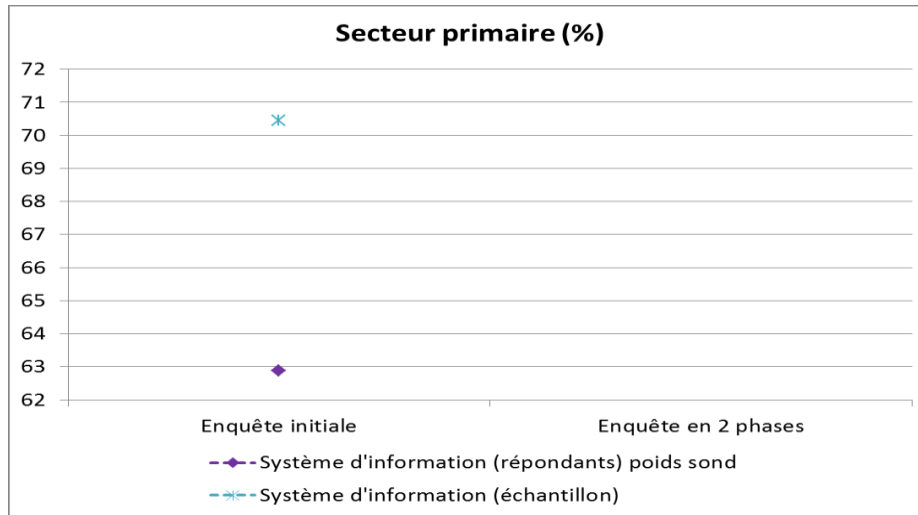
- **Secteur d'activité primaire**
- **Pour les salariés : contrat de travail en CDI**
- Statut d'emploi : salarié
- Pour les non-salariés : surface agricole utile en ares

Disposer de données gold standard

→ Etude restreinte aux personnes ayant eu un seul emploi affilié à la MSA en 2010 dans les bases de la MSA (n=7484)

Résultats

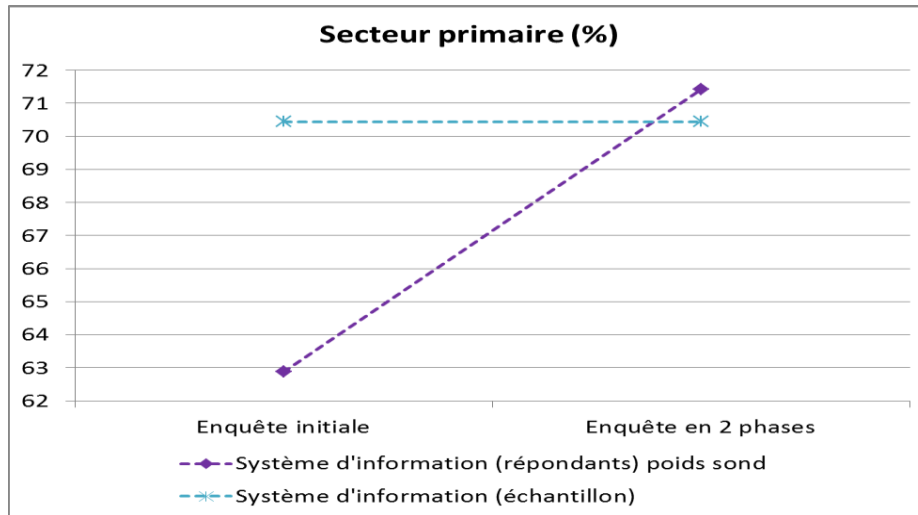
Sans correction de la non-réponse



Erreur de non-réponse

Résultats

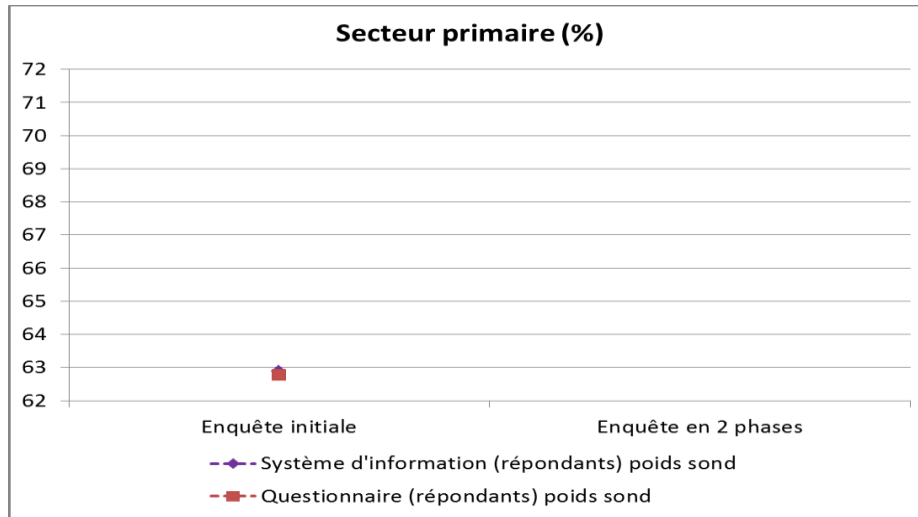
Sans correction de la non-réponse



Erreur de non-réponse

Résultats

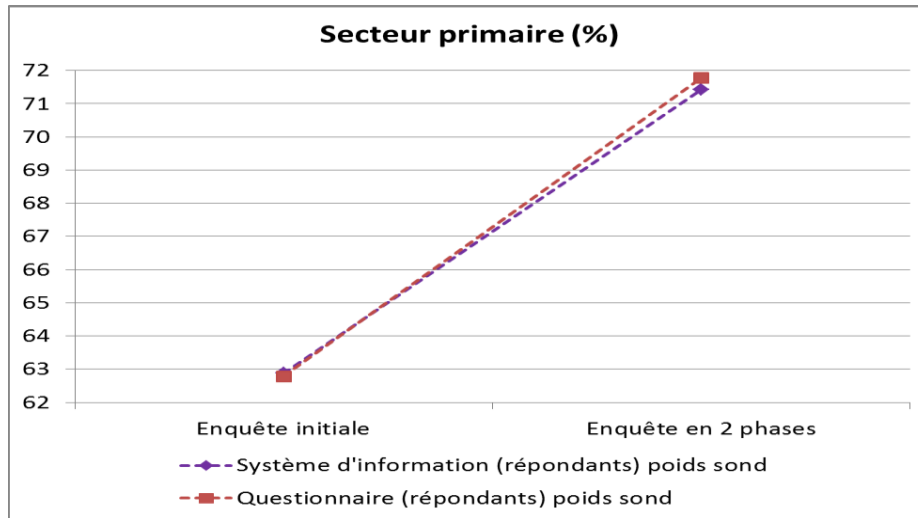
Sans correction de la non-réponse



Erreur de mesure

Résultats

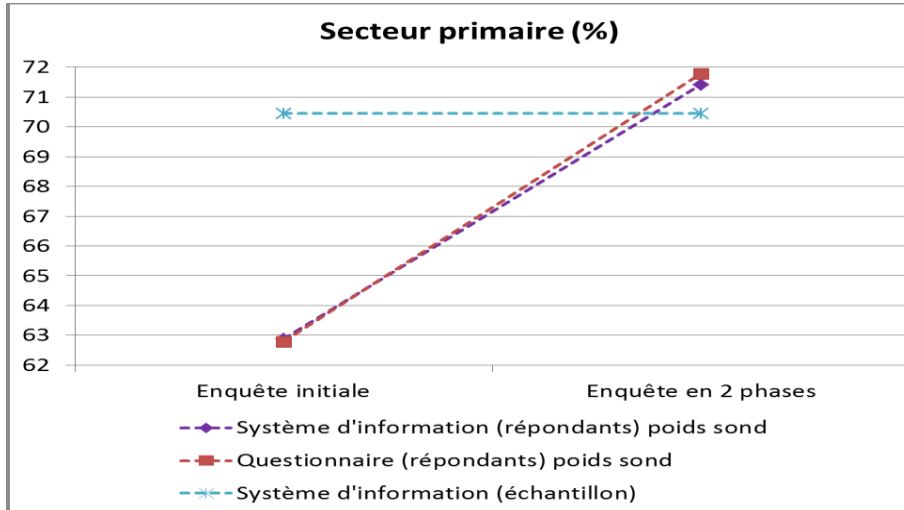
Sans correction de la non-réponse



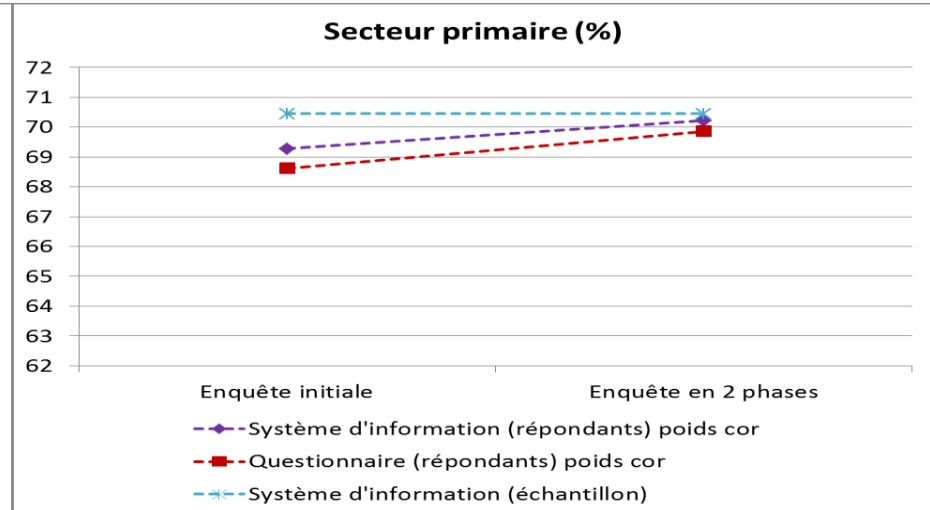
Erreur de mesure

Résultats

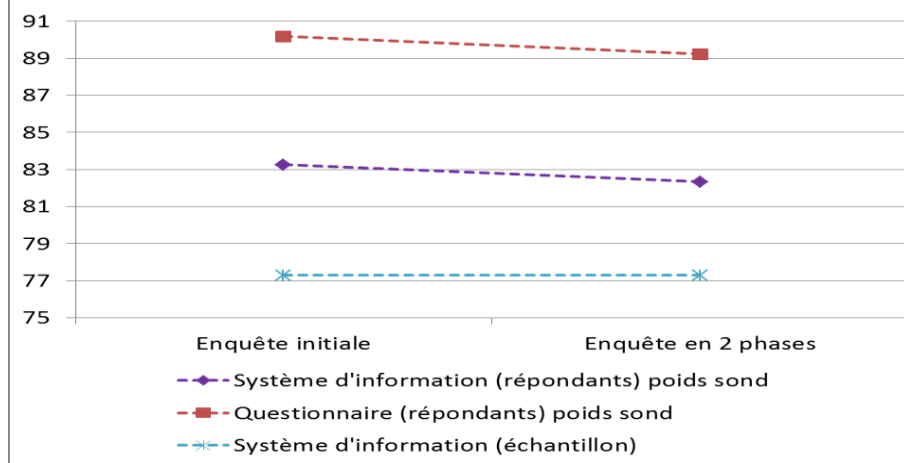
Sans correction de la non-réponse



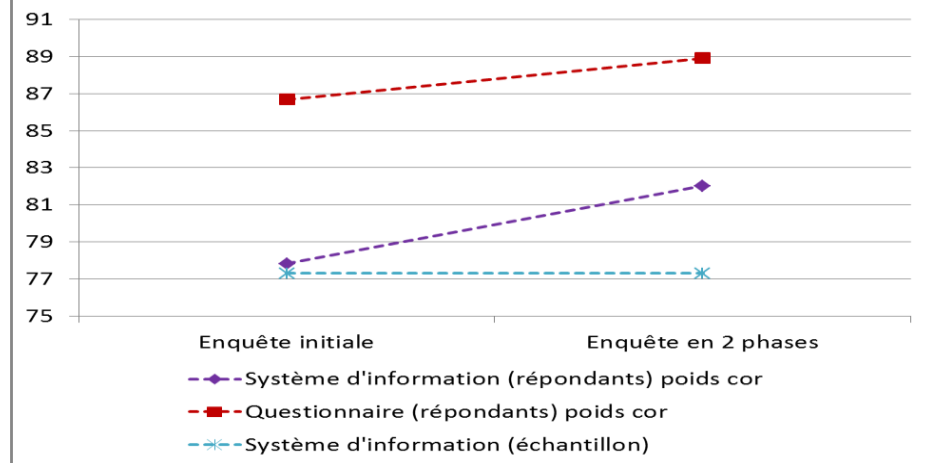
Avec correction de la non-réponse



Contrat CDI (salarié) (%)



Contrat CDI (salarié) (%)



Forces et faiblesses

- **Faiblesses**

- Erreur de mesure / erreur de non-réponse : sur quelques variables relatives à l'emploi
- Sélection des emplois les plus stables

- **Forces**

- Données du SNIIRAM et de la MSA appariées sur l'échantillon tiré au sort
 - Correction de la non-réponse, prévalences gold-standard (...)
- Protocole original qui a permis d'étudier la non-réponse sous plusieurs angles

Discussion générale

- **Comparaison avant/après correction de la NR :**
 - Erreur de non-réponse (ENR) nettement diminuée → on corrige bien !
 - Erreur de mesure (EM) stable ou plus élevée après correction de la non-réponse
 - **Après correction de la non-réponse :**
 - Prévalences à l'enquête initiale équivalentes à celles d'enquête en deux phases
 - Contribution de l'EM nettement plus élevée que celle de l'ENR (sauf pour secteur primaire)
- Intérêt d'étudier le compromis entre ENR et EM après correction de la non-réponse

Conclusions - Perspectives

- **Enquête initiale avec appariement des données du SNIIRAM et de la MSA suffisante pour les variables étudiées**
- **Travaux complémentaires :**
 - Utilisation d'autres définitions de difficulté à joindre les personnes
- **Intérêt ++ de diversifier les sources de données dans une enquête**

MERCI POUR VOTRE ATTENTION