

Biais de sélection à l'inclusion dans une cohorte de surveillance des risques professionnels

Gaëlle Santin¹, Laetitia Bénézet¹, Béatrice Geoffroy¹, Pauline Delézire¹, Juliette Chatelot¹, Rémi Sitta², Jean Bouyer³, Alice Gueguen²
Groupe cohortes-SniirAM^{1,2,4}, Equipe Coset¹

¹ *Institut de veille sanitaire, Département santé travail, Saint-Maurice*

² *CESP - Inserm 1018, Plateforme de recherche « Cohortes en population », Villejuif*

³ *CESP - Inserm 1018, Epidémiologie de la reproduction et du développement de l'enfant, Le Kremlin-Bicêtre*

⁴ *Institut de veille sanitaire, Direction scientifique, Saint-Maurice*



Contexte

- Taux de non-réponse de plus en plus élevé dans les enquêtes → prise en compte de la non-réponse *a posteriori* de plus en plus fondamentale
- Difficultés d'accès à des informations pertinentes sur l'ensemble de l'échantillon
→ Une façon de procéder : apparier l'échantillon tiré au sort avec des données issues de systèmes d'informations existants
- Dans les études épidémiologiques ayant eu recours à des informations issues de systèmes d'information existants pour étudier la non-réponse, non-réponse associée à des :
 - Variables socio-démographiques
 - Variables socio-professionnelles
 - Variables de santé
- Très peu d'études épidémiologiques prenant en compte toutes ces variables pour étudier la non-réponse ou utilisant ces informations pour corriger la non-réponse



Objectifs

- Etudier les facteurs prédictifs de la non-réponse dans une enquête de surveillance épidémiologique des risques professionnels (cohorte pilote Coset-MSA à l'inclusion) en utilisant des données administratives relatives à la santé (SNIIR-AM) et à l'activité professionnelle (MSA)
- Estimer des prévalences après repondération en prenant en compte les facteurs liés à la non-réponse



Population et méthodes (1)

Phase pilote Coset-MSA – Inclusion

Cohorte pour la Surveillance Épidémiologique en lien avec le Travail
chez les travailleurs affiliés à la Mutualité Sociale Agricole

Population d'étude	Salariés et non salariés agricoles ayant travaillé au moins 90 jours en 2008 dans un des 5 départements pilotes
Echantillonnage	Tirage au sort de 10 000 personnes stratifié sur le sexe, l'âge, le statut (salarié/non salarié), le département
Période d'enquête	Février-mai 2010
Mode de recueil des données	<u>Actif</u> : autoquestionnaire postal (AQ) <u>Passif</u> : systèmes d'information existants (SI)



Population et méthodes (2)

Phase pilote Coset-MSA – Inclusion

Cohorte pour la Surveillance Épidémiologique en lien avec le Travail
chez les travailleurs affiliés à la Mutualité Sociale Agricole

**Données disponibles chez
les répondants et les
non-répondants (SI)**

Variables de stratification (2008) : sexe, âge, département, statut

SNIIR-AM :

- *Remboursement de soins (2008-2010)* :
nombre de recours à un professionnel de santé, par type
nombre de boîtes de médicaments, par type (classification ATC)
montant total des prestations

- *Hospitalisations via le PMSI (2008-2009)* : oui/non

MSA : *dernier emploi principal en date (2008 ou 2009)*

statut, secteur d'activité, durée d'emploi, nombre d'emplois salariés dans l'année

**Données disponibles
uniquement chez les
répondants (AQ)**

Santé perçue, échelles de santé (TMS, symptômes dépressifs, asthme), consommation de tabac et d'alcool, historiques d'emplois et d'expositions professionnelles



Population et méthodes (3)

Etude des facteurs liés à la réponse

- Sélection des facteurs par régression logistique (seuil 5%) :
 - Etape 1 : séparément dans chaque source de données (strates, SNIIR-AM, MSA)
 - Etape 2 : en combinant les facteurs sélectionnés par les trois modèles précédents

Correction de la non-réponse par repondération

- Méthode des scores [Little 1986; Eltinge&Yansaneh 1997; Haziza&Beaumont 2007]

Estimation des prévalences

- Utilisation de pondérations pour l'estimation de prévalence de :
 - Variables issues des systèmes d'information (→ chez les répondants et dans l'échantillon complet) « nombre d'actes » « secteur d'activité » + Calcul de l'erreur relative
 - Variables issues du questionnaire (→ chez les répondants) « santé perçue » « j'ai le respect que je mérite au travail » « catégorie sociale »



Résultats

Taux d'appariement échantillon TAS / données SI

- ~ 650 personnes sans tentative d'appariement
 - ~ 120 refus exprimés,
 - ~ 520 plis *a priori* non distribués
- Taux d'appariement données professionnelles MSA :
98% (n=9364)
- Appariement données SNIIR-AM :

	n	2008 (%)	2009 (%)	2010 (%)
Remboursement de soins ambulatoires	9358	89,0	89,5	88,2
PMSI	9358	10,7	11,3	-

→ Par an, ~10% de personnes sans remboursement de soins et ~ 90% de personnes non hospitalisées



Résultats

Taux de participation

Participation selon les strates et les données **SNIR-AM** et **MSA**

- Taux de réponse moyen : 24%
- Participation selon les variables de stratification, du **SNIR-AM** et de la **MSA**

		Multivarié (*)			
		OR	IC 95%		p
Age	18 à 34 ans	1,0			0,001
	35 à 49 ans	1,2	1,1	1,4	
	50 à 65 ans	1,3	1,1	1,5	
Caisse d'affiliation MSA	Bouches du Rhône	1,0			<0,0001
	Pyrénées Atlantiques	1,1	0,9	1,3	
	Finistère	1,1	1,0	1,4	
	Pas de Calais	1,4	1,2	1,6	
	Saône et Loire	1,5	1,3	1,8	

(*) *Modèle final issu des données combinées « strates » « SNIR-AM » « MSA » - variables avec $p < 0,05$*



Résultats

Participation selon les strates et les données SNIIR-AM et MSA

Entre 2008 et 2010		Multivarié (*)			
		OR	IC 95%		p
Recours à un médecin généraliste	Aucun	1,0			0,008
	Entre 1 et 4 fois	1,5	1,1	2,0	
	Entre 5 et 9 fois	1,6	1,2	2,2	
	Au moins 10 fois	1,4	1,0	2,0	
Recours à un médecin spécialiste	Aucun	1,0			<0,0001
	Entre 1 et 4 fois	1,1	0,9	1,2	
	Entre 5 et 9 fois	1,2	1,0	1,4	
	Au moins 10 fois	1,6	1,3	1,9	
Recours à un dentiste	Aucun	1,0			<0,0001
	Entre 1 et 9 fois	1,3	1,2	1,5	
	Au moins 10 fois	1,2	1,0	1,4	
Recours à un laboratoire d'analyses	Aucun	1,0			0,0005
	Entre 1 et 9 fois	1,3	1,1	1,5	
	Au moins 10 fois	1,4	1,2	1,7	
Recours à un auxiliaire médical	Aucun	1,0			0,01
	Entre 1 et 9 fois	1,0	0,9	1,1	
	Au moins 10 fois	1,2	1,0	1,4	
Recours à un fournisseur "autre" (prothèses, lunettes...)	Aucun	1,0			0,02
	Au moins 1 fois	1,1	1,0	1,3	
Nombre total de boîtes de médicaments facturées	Aucun	1,0			0,009
	Entre 1 et 199	0,8	0,6	1,1	
	Entre 200 et 799	0,7	0,5	1,0	
	Au moins 800	0,6	0,4	0,8	
Hospitalisation	Aucune	1,0			0,04
	Au moins 1	0,9	0,8	1,0	

9 (*) *Modèle final issu des données combinées « strates » « SNIIR-AM » « MSA » - variables avec p<0,05*



Résultats

Participation selon les strates et les données SNIIR-AM et MSA

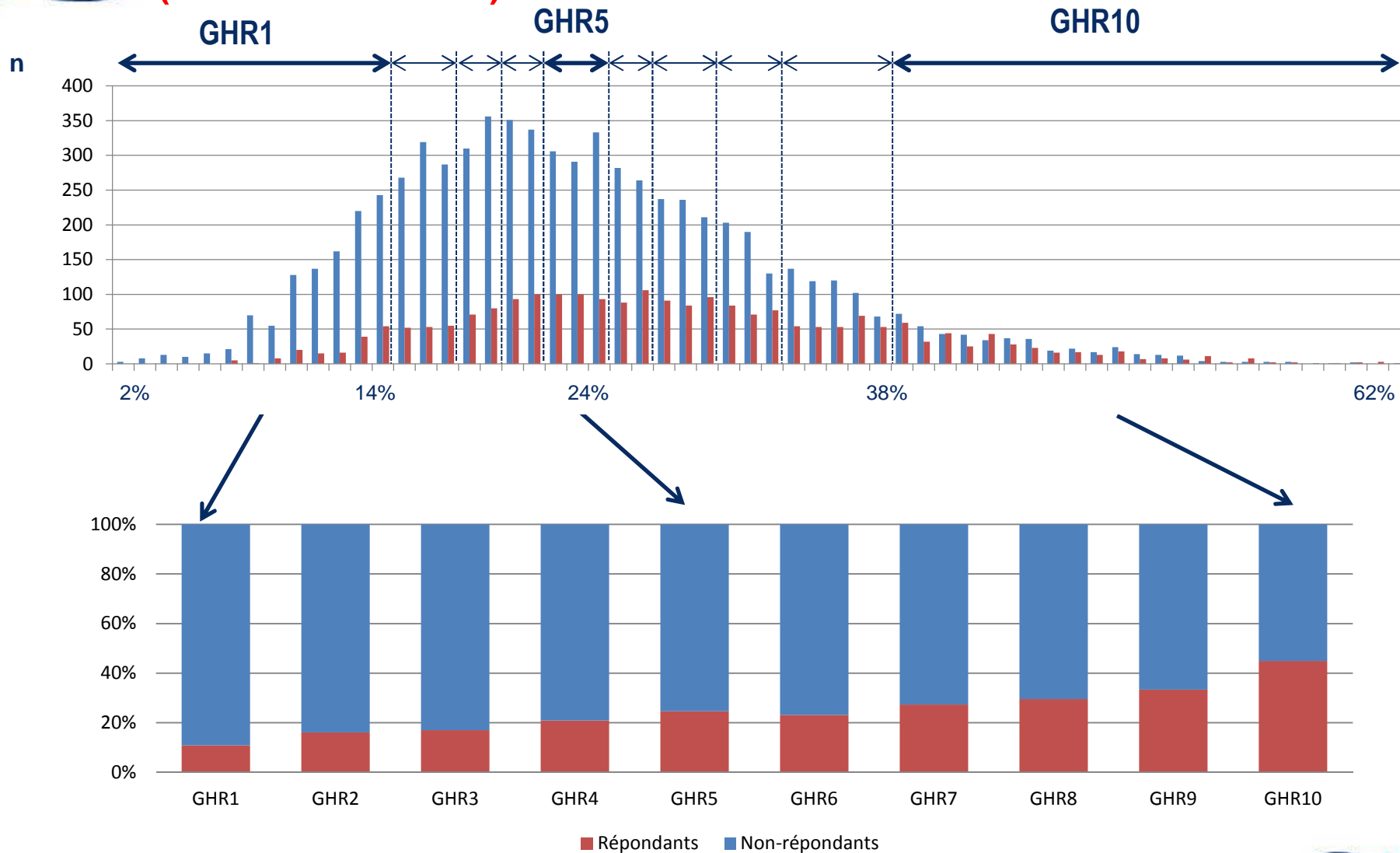
Dernier emploi principal (en 2008 ou 2009)		Multivarié (*)			p
		OR	IC 95%		
Statut du dernier emploi principal	Non salarié	1,0			<0,0001
	Salarié	1,4	1,2	1,7	
Année du dernier emploi principal	2008	1,0			0,01
	2009	1,4	1,1	1,7	
Secteur d'activité MSA	Cultures spécialisées	1,0			<0,0001
	Viticulture	0,9	0,7	1,2	
	Elevage spécialisé gros animaux	1,0	0,9	1,2	
	Elevage spécialisé petits animaux	1,1	0,8	1,5	
	Polyculture, polyélevage	0,9	0,7	1,0	
	Haras...	0,7	0,4	1,1	
	Travaux forestiers	0,9	0,6	1,3	
	Entreprises de travaux agricoles	0,7	0,6	0,9	
	Artisanat rural	0,2	0,1	0,6	
	Coopération	0,9	0,8	1,1	
	Organismes de service	1,3	1,1	1,6	
Activités diverses	1,1	0,7	1,6		
Durée du dernier emploi principal	<6 mois	1,0			0,002
	[6 mois-1 an[1,3	1,0	1,8	
	[1-6[ans	1,4	1,1	1,9	
	[6-10[ans	1,2	0,9	1,6	
	[10-20[ans	1,6	1,2	2,1	
	>=20 ans	1,5	1,1	2,0	

(*) *Modèle final issu des données combinées « strates » « SNIIR-AM » « MSA » - variables avec p<0,05*



Résultats

Construction des groupes homogènes de réponse (méthode des scores)





Résultats

Prévalence de variables issues des systèmes d'information

		MCAR
		%
Nombre d'actes	Aucun	2,9
	1_49	43,5
	50_499	27,9
	>=500	25,8
Nombre d'actes (moyen)		75,7
Secteur d'activité	Primaire	57,2
	Secondaire	6,2
	Tertiaire	36,6



Résultats

Prévalence de variables issues des systèmes d'information

		MCAR	MAR/strates (MAR1)
		%	%
Nombre d'actes	Aucun	2,9	2,8
	1_49	43,5	45,6
	50_499	27,9	27,2
	>=500	25,8	24,4
Nombre d'actes (moyen)		75,7	73,5
Secteur d'activité	Primaire	57,2	58,8
	Secondaire	6,2	5,6
	Tertiaire	36,6	35,5



Résultats

Prévalence de variables issues des systèmes d'information

		MCAR	MAR/strates (MAR1)	MAR/strates, SNIIR-AM, MSA (MAR2)
		%	%	%
Nombre d'actes	Aucun	2,9	2,8	4,3
	1_49	43,5	45,6	49,1
	50_499	27,9	27,2	25,1
	>=500	25,8	24,4	21,5
Nombre d'actes (moyen)		75,7	73,5	66,5
Secteur d'activité	Primaire	57,2	58,8	61,8
	Secondaire	6,2	5,6	5,9
	Tertiaire	36,6	35,5	32,3



Résultats

Prévalence de variables issues des systèmes d'information

		MCAR	MAR/strates (MAR1)	MAR/strates, SNIIR-AM, MSA (MAR2)	Echantillon complet
		%	%	%	%
Nombre d'actes	Aucun	2,9	2,8	4,3	4,4
	1_49	43,5	45,6	49,1	49,6
	50_499	27,9	27,2	25,1	26,1
	>=500	25,8	24,4	21,5	19,9
Nombre d'actes (moyen)		75,7	73,5	66,5	64,7
Secteur d'activité	Primaire	57,2	58,8	61,8	61,2
	Secondaire	6,2	5,6	5,9	6,5
	Tertiaire	36,6	35,5	32,3	32,3



Résultats

Prévalence de variables issues des systèmes d'information

		MCAR	MAR/strates (MAR1)	MAR/strates, SNIIR-AM, MSA (MAR2)	Echantillon complet	Erreur relative		
		%	%	%	%	MCAR / Complet	MAR1 / Complet	MAR2 / Complet
Nombre d'actes	Aucun	2,9	2,8	4,3	4,4	-35,2	-37,3	-2,1
	1_49	43,5	45,6	49,1	49,6	-12,3	-7,9	-1,0
	50_499	27,9	27,2	25,1	26,1	6,9	4,4	-3,7
	>=500	25,8	24,4	21,5	19,9	29,4	22,3	7,9
Nombre d'actes (moyen)		75,7	73,5	66,5	64,7	17,0	13,5	2,7
Secteur d'activité	Primaire	57,2	58,8	61,8	61,2	-6,5	-3,8	1,1
	Secondaire	6,2	5,6	5,9	6,5	-4,8	-13,7	-9,8
	Tertiaire	36,6	35,5	32,3	32,3	13,2	10,0	-0,1



Résultats

Prévalence de quelques variables de l'autoquestionnaire

	Sans correction de la non-réponse (MCAR)		
	%	IC 95%	
Etat de santé général perçu			
Très bon	55,8	53,8	57,9
Pas/Moyennement bon	44,2	42,1	46,2
Catégorie sociale			
Agriculteur exploitant	27,8	25,9	29,7
Artisans commerçants chef d'entreprise	2,0	1,4	2,6
Cadre, profession intellectuelle supérieure	7,5	6,4	8,7
Profession intermédiaire	17,0	15,3	18,6
Employé	14,0	12,5	15,5
Ouvrier	31,7	29,7	33,8
J'ai le respect que je mérite au travail			
D'accord	59,4	57,2	61,7
Pas d'accord, ça ne me perturbe pas du tout	13,4	11,8	14,9
Pas d'accord, ça me perturbe un peu	16,5	14,8	18,3
Pas d'accord, ça me perturbe	6,8	5,6	8,0
Pas d'accord, ça me perturbe beaucoup	3,8	3,0	4,7



Résultats

Prévalence de quelques variables de l'autoquestionnaire

	Sans correction de la non-réponse (MCAR)			Après correction de la non-réponse (MAR1) (strates)		
	%	IC 95%		%	IC 95%	
Etat de santé général perçu						
Très bon	55,8	53,8	57,9	57,0	54,8	59,1
Pas/Moyennement bon	44,2	42,1	46,2	43,0	40,9	45,2
Catégorie sociale						
Agriculteur exploitant	27,8	25,9	29,7	29,7	27,6	31,7
Artisans commerçants chef d'entreprise	2,0	1,4	2,6	2,3	1,6	3,1
Cadre, profession intellectuelle supérieure	7,5	6,4	8,7	7,1	6,0	8,3
Profession intermédiaire	17,0	15,3	18,6	16,6	15,0	18,3
Employé	14,0	12,5	15,5	12,6	11,2	14,0
Ouvrier	31,7	29,7	33,8	31,6	29,5	33,7
J'ai le respect que je mérite au travail						
D'accord	59,4	57,2	61,7	59,4	57,0	61,8
Pas d'accord, ça ne me perturbe pas du tout	13,4	11,8	14,9	13,5	11,8	15,1
Pas d'accord, ça me perturbe un peu	16,5	14,8	18,3	16,5	14,7	18,3
Pas d'accord, ça me perturbe	6,8	5,6	8,0	6,6	5,5	7,8
Pas d'accord, ça me perturbe beaucoup	3,8	3,0	4,7	3,9	3,0	4,9



Résultats

Prévalence de quelques variables de l'autoquestionnaire

	Sans correction de la non-réponse (MCAR)			Après correction de la non-réponse (MAR1) (strates)			Après correction de la non-réponse (MAR2) (strates, SNIIR-AM, MSA)		
	%	IC 95%		%	IC 95%		%	IC 95%	
Etat de santé général perçu									
Très bon	55,8	53,8	57,9	57,0	54,8	59,1	58,5	56,2	60,7
Pas/Moyennement bon	44,2	42,1	46,2	43,0	40,9	45,2	41,5	39,3	43,8
Catégorie sociale									
Agriculteur exploitant	27,8	25,9	29,7	29,7	27,6	31,7	30,8	28,6	33,0
Artisans commerçants chef d'entreprise	2,0	1,4	2,6	2,3	1,6	3,1	3,1	2,1	4,1
Cadre, profession intellectuelle supérieure	7,5	6,4	8,7	7,1	6,0	8,3	6,0	5,0	7,0
Profession intermédiaire	17,0	15,3	18,6	16,6	15,0	18,3	14,5	12,9	16,1
Employé	14,0	12,5	15,5	12,6	11,2	14,0	11,4	10,1	12,8
Ouvrier	31,7	29,7	33,8	31,6	29,5	33,7	34,1	31,8	36,5
J'ai le respect que je mérite au travail									
D'accord	59,4	57,2	61,7	59,4	57,0	61,8	60,3	57,8	62,7
Pas d'accord, ça ne me perturbe pas du tout	13,4	11,8	14,9	13,5	11,8	15,1	13,5	11,8	15,3
Pas d'accord, ça me perturbe un peu	16,5	14,8	18,3	16,5	14,7	18,3	16,3	14,4	18,2
Pas d'accord, ça me perturbe	6,8	5,6	8,0	6,6	5,5	7,8	6,2	5,1	7,4
Pas d'accord, ça me perturbe beaucoup	3,8	3,0	4,7	3,9	3,0	4,9	3,7	2,7	4,6



Synthèse des résultats

1. Etude des facteurs liés à la non-réponse

- Données de santé : propension à répondre :
 - plus élevée chez les personnes intéressées par leur santé ? (recours aux soins)
 - moins élevée chez les personnes ayant un mauvais état de santé ? (consommation de soins)
- Données professionnelles : propension à répondre plus élevée :
 - chez les actifs les plus disponibles ? (salarié/non salarié ; secteur services/agricoles)
 - chez les personnes les moins « précaires » ? (durée dans l'emploi > 6 mois)

2. Facteurs d'ajustement de la non-réponse

- Taux de réponses dans les GHR obtenus par la méthode des scores : de 10% à 45%
1 et 2 → Choix des variables SNIIR-AM et MSA pertinent

3. Enquête de surveillance épidémiologique des risques professionnels

- Objectifs descriptifs
- Non-réponse liée à des variables relatives à la santé et au travail
→ Biais de sélection potentiels → Intérêt +++ des données SNIIR-AM et MSA

4. Impact sur les estimations de prévalence

- En général plutôt modéré mais existant, différent selon les variables



Perspectives

- Modélisation :
 - Modèles différents pour les salariés et les non-salariés
 - Variables quantitatives en continu (splines)
- Données supplémentaires : indemnités journalières, détail des hospitalisations...
 - Hypothèse MAR plus vraisemblable
- Exploitation des données de l'enquête complémentaire réalisée auprès d'un échantillon de 500 non-participants (taux de réponse : 65%)
 - Génération d'hypothèse NMAR (analyse de sensibilité)
- Estimation correcte de la variance



Remerciements

MSA : Florian Brémaud, Yves Cosset, Alain Pelc, Nicolas Viarouge

CnamTS : David Diaz, Laurent Duchet, Philippe Ricordeau

Equipe Constances : Marcel Goldberg, Julie Gourmelen, Marie Zins

DiSQ-InVS : Javier Nicolau

DST-InVS : Jean-Luc Marchand