

LES COHORTES ÉPIDÉMIOLOGIQUES PROSPECTIVES

Principes – Difficultés – Solutions

Marcel Goldberg¹ & Marie Zins^{1,2}

1 : Unité 687 Inserm-Cnamts

2 : Équipe Risques Post Professionnels - Cohortes du
Cetaf

Objectifs généraux et principes des cohortes

Cohorte : enquête épidémiologique avec suivi individuel longitudinal d'un groupe de sujets : prise en compte du temps à l'échelle individuelle

- 1. Étude de l'évolution d'un phénomène populationnel dans le temps** (la fréquence du tabagisme diminue-t-elle ? L'incidence des cancers augmente-t-elle ?)
 - Approches pertinentes : enquête transversale répétée ou enregistrement systématique permanent (mortalité, registres du cancer).
- 2. Évolution d'un processus à l'échelle individuelle**
 - « Histoire naturelle » des processus physiopathologiques et des maladies
 - Causalité : l'exposition à un « facteur de risque » modifie-t-elle le risque de survenue d'une maladie ?



Nécessité d'analyser la temporalité à l'échelle individuelle : études de cohorte.

L'apport des cohortes prospectives

- Études transversales : limites connues pour évaluer des évolutions temporelles et des interactions de différents facteurs
 - État de santé
 - Caractéristiques personnelles
 - Conditions de vie
 - Environnement
- Études rétrospectives : difficultés de la mesure des expositions (biais, imprécisions...)

- Les cohortes prospectives
 - Expositions mesurées avant la survenue des effets analysés
 - Problèmes de santé étudiés en termes d'états précliniques, d'incidence, de mortalité
 - Analyse longitudinale : prise en compte du temps

Apports des cohortes prospectives

□ Scientifique

- Connaissance de l'histoire naturelle (maladies, expositions...)
- Meilleures conditions méthodologiques pour établir des relations de causalité

□ Santé publique (selon représentativité de l'échantillon)

- Analyse des trajectoires médicale, professionnelle et sociale des personnes
- Outil de surveillance épidémiologique
- Outil d'évaluation d'intervention

Possibilité d'adapter les objectifs au cours du suivi

« Une cohorte doit être conçue pour répondre à plusieurs questions de recherche épidémiologique, clinique, biologique ou de santé publique même si certaines ne sont pas encore formulées de façon précise au démarrage de la cohorte »

(ANRS)

Cohortes « spécialisées » et cohortes « généralistes »

□ Cohortes spécialisées

- centrées sur un problème spécifique (pathologie et/ou groupe de population)
- sujets sélectionnés selon des caractéristiques particulières
- effectif relativement restreint
- données très détaillées avec investigations biocliniques approfondies.

□ Cohortes généralistes

- en population générale
- de grande taille
- couverture large de problèmes de santé et de déterminants
- données relativement superficielles
- ouverture vers des utilisateurs diversifiés
- cohortes « mixtes »

Cohortes généralistes : « plateformes scientifiques » permettant d'étudier de nombreuses questions

Les enjeux actuels des cohortes épidémiologiques

- Effectifs de plus en plus importants
- Coûts élevés
- Suivi à très long terme

Quelques grandes cohortes

❑ À l'étranger

- *One Million Women Study* (GB) : > 1 million femmes
- *UK Biobank* (GB) : 500 000 personnes
- *General Practice Research Data Base* (GB) : 4 millions personnes
- *Norwegian Mother and Child Cohort Study* (N) : 100 000 femmes, 100 000 enfants, 70 000 pères
- *Nurses'Health Study* (USA) : 239 000 femmes

❑ En France

- E3N : 100 000 femmes adhérentes MGEN
- Suvimax : 13 000 "tout venant"
- GAZEL : 20 000 agents EDF-GDF

Durée de suivi des cohortes

□ À l'étranger

- *National Survey of Health and Development* (GB) : enfants nés en 1946
- *Framingham* (USA) : 1948
- *British Doctors' Study* (GB) : 1951

□ En France

- *Étude Prospective Parisienne* : 1967
- *PAQUID* : 1988
- *GAZEL* : 1989

En France

- Cohortes relativement récentes
- Effectif restreint

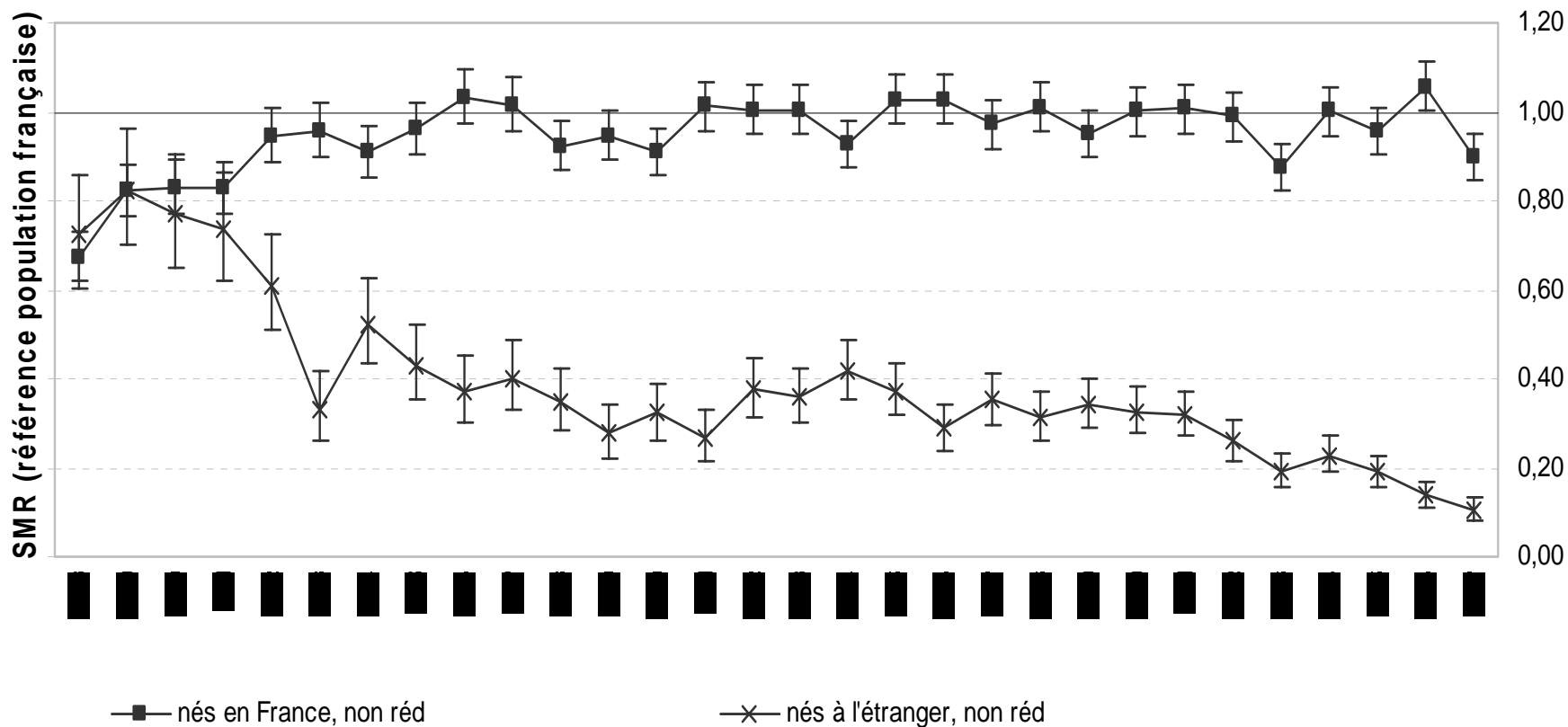
Problèmes de fonctionnement

- ❑ Nombre trop faible d'épidémiologistes
- ❑ Difficultés d'ordre financier : coût d'une cohorte prospective en population : 25 € à 250 €/an/ sujet (hors examens spécialisés)
- ❑ Règles de gestion des organismes publics
- ❑ Implication à long terme des équipes
- ❑ Manque de personnel technique stable : absence de statut reconnu pour ce type d'activité ; durée des projets incompatible avec un trop fort renouvellement (continuité des procédures et des recueils de données)
- ❑ Sous utilisation des bases de données alimentées par les organismes de protection sociale et médicale

Aspects techniques de la gestion des cohortes

- ❑ Transmission et gestion sécurisée des bases de données : problèmes « classiques » dans le domaine des données de santé
- ❑ Extraction de données concernant des sujets sélectionnés pour le suivi des sujets et l'identification des événements d'intérêt
- ❑ Principales activités
 - « Traçage » de sujets inclus
 - Recueil de données à différentes sources
 - Contrôle et validation des données
 - Appariement de données individuelles en provenance de différentes sources

Hommes



Source: Geoffroy-Perez, Étude COSMOP, 2005

Les bases de données nationales d'origine administrative : un atout potentiel

- ❑ « Modèle scandinave » : très grandes études par appariement de bases de données nationales

- ❑ Extraction de données concernant des sujets sélectionnés : recours aux soins, traitements, pathologies, parcours social et professionnel

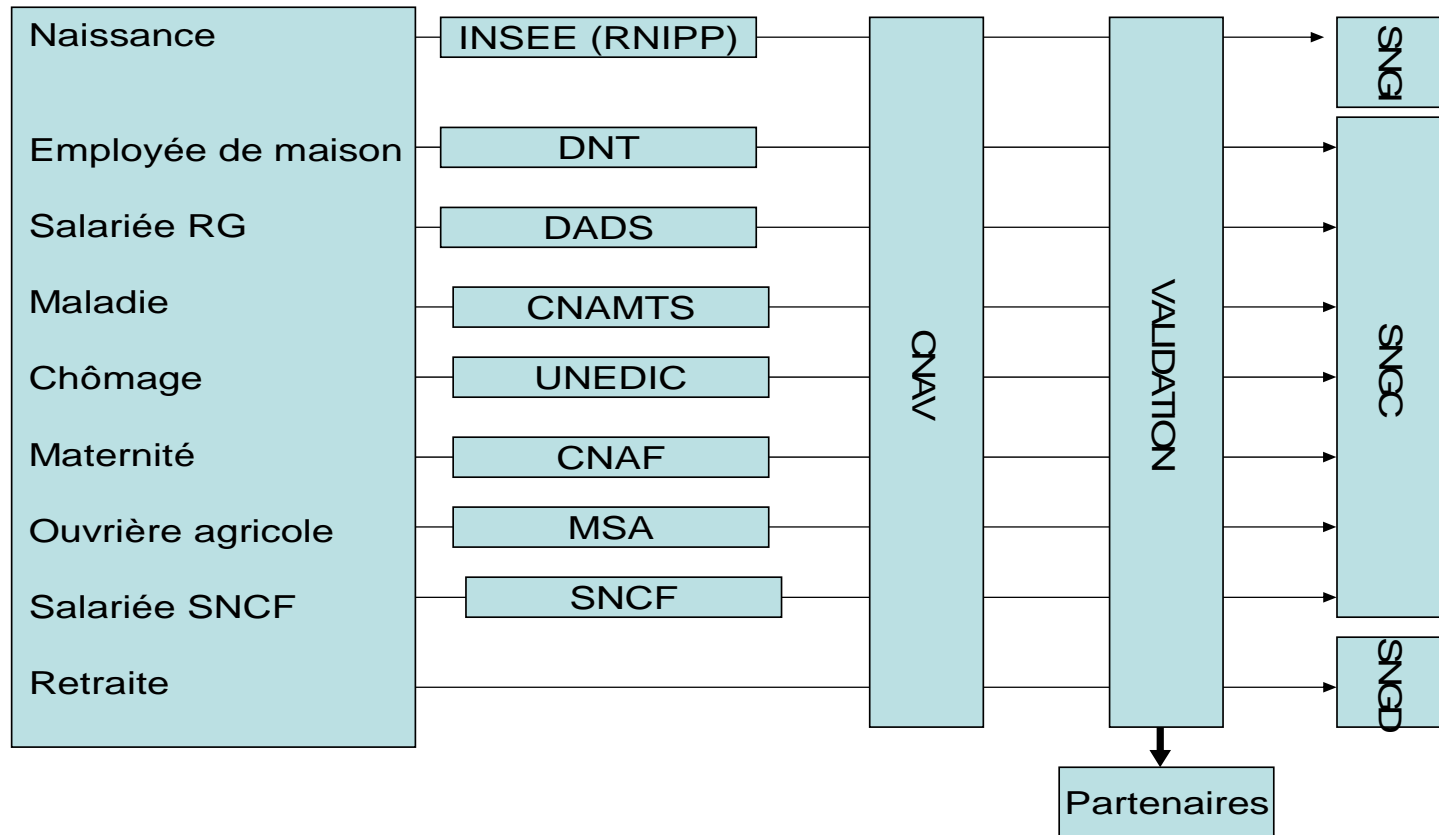
- ❑ Deux grands systèmes d'informations de nature médicale et administrative contiennent des données individuelles exhaustives d'intérêt potentiel pour les épidémiologistes dont peu de pays disposent à l'échelle nationale
 - Bases de données de la Cnav
 - Bases de données de l'AM

Traçage de sujets inclus

Environ 3 millions de foyers (5,6 millions d'individus) déménagent chaque année en France

- Caisses d'allocations familiales : versent les prestations sociales ;
- Sujets salariés : Déclarations annuelles des données sociales (DADS) qui contiennent l'adresse connue par les employeurs ;
- Organismes servant les prestations d'assurance maladie ;
- Retraités : CRAM qui assurent le versement des pensions de vieillesse
- La Poste

CNAV : données socioprofessionnelles*



*Ribet C et al : Les bases de données socioprofessionnelles. (<http://www.rppc.fr/publications.html>)

Les bases de données de l'AM

- ❑ Données à vocation médico-tarifaire regroupées dans le SNIIR-AM
 - Consommation de soins remboursés avec codes détaillés : prescripteurs, exécutants, actes, médicaments
 - ALD
 - AT-MP
 - PMSI
- ❑ Étude préliminaire*
 - Bases a vocation médico-tarifaire
 - Très riche en informations (exhaustivité ?)
 - Mais lourd travail de validation (données médicales)
- ❑ Accès par tiers de confiance

* Mireille Cœuret-Pellicer et al : Les bases de données del'Assurance Maladie. (<http://www.rppc.fr/publications.html>)

Avantages

- ❑ Données médicalisées, structurées et codées de manière standardisée
- ❑ Données parfois plus fiables que celles obtenues par déclaration
- ❑ Quasi exhaustivité de la population cible : absence de biais de sélection ; effectifs immenses
- ❑ Quasi absence de perdus de vue pendant le suivi

Limites : problèmes de validité dans une optique épidémiologique

- ❑ Remboursements : pas d'information sur les maladies traitées, pas l'automédication et les prestations non présentées au remboursement
- ❑ ALD : imprécision des diagnostics (qualité du codage non évaluée), absence d'exhaustivité des cas déclarés, risque de double déclaration
- ❑ PMSI : diagnostic codé insuffisant ; nécessité de développer des algorithmes complexes alliant les codes diagnostic aux codes actes spécifiques de la pathologie étudiée
- Procédures de validation : retour au médecin traitant, confrontation avec les déclarations des sujets, croisement avec d'autres sources (données de registre, causes de décès...)

Deux exemples

1. **GAZEL** : une cohorte pour la recherche (www.gazel.inserm.fr)
2. **Constances** : une cohorte pour la recherche et l'information en santé publique (www.constances.fr)

La cohorte GAZEL : un laboratoire épidémiologique ouvert

Électricité de France - Gaz de France

- ❑ Environ 140 000 agents répartis sur tout le territoire : large éventail de catégories socio-économiques et de métiers, très stable (toute la carrière dans l'entreprise, régime particulier de retraite)
- ❑ Médecine du travail et Sécurité Sociale intégrées à l'entreprise : disponibilité de données de santé
- ❑ Fichiers de gestion du personnel : disponibilité de nombreuses données socio-économiques et professionnelles

GAZEL : un laboratoire epidemiologique ouvert



Suivi à long terme

- Recueil systématique et prospectif de données variées
- Accès à la base de données et aux adresses des volontaires dans le cadre de recherches spécifiques

Projets de recherche sur des problèmes de santé variés et sur les facteurs liés à la santé

La cohorte à l'inclusion

□ Janvier 1989



- invitation à participer à une enquête épidémiologique pour aider la recherche médicale
- population cible : hommes entre 40 et 50 ans ; femmes entre 35 et 50 ans

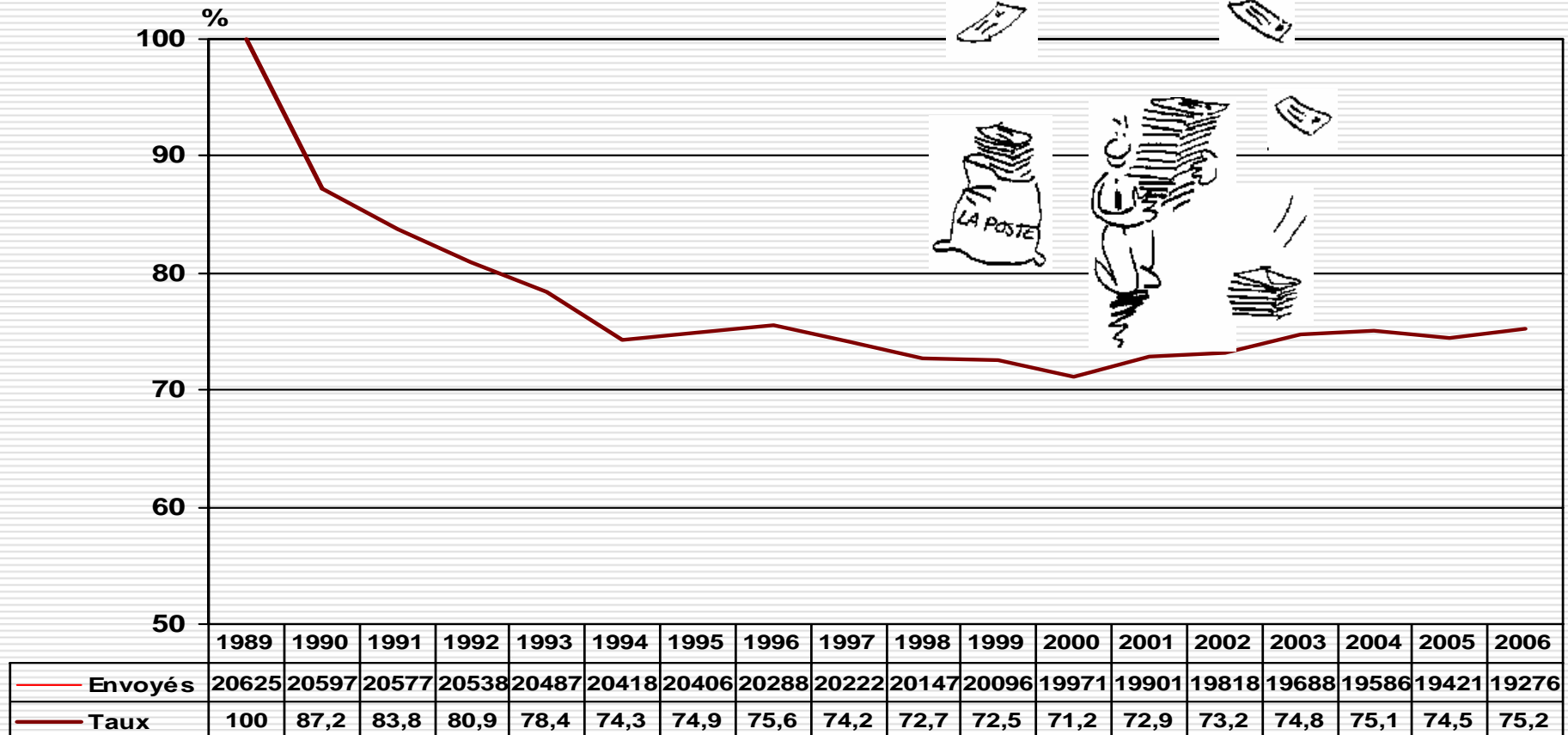
□ Taux de participation : 44 , 5 %

- 20 625 volontaires : 15 011 hommes et 5 614 femmes

Recueil de données

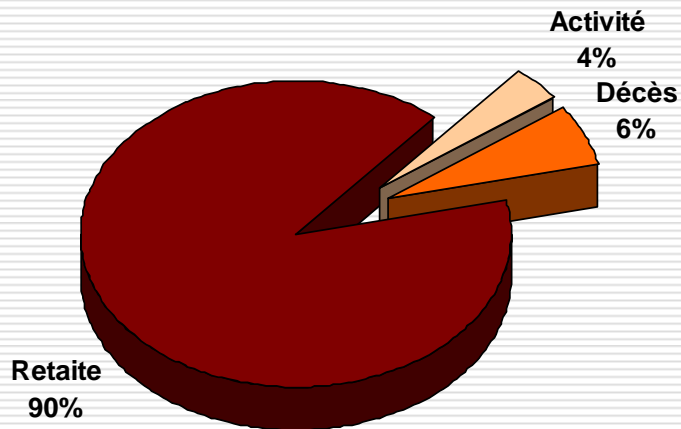
Données par type de suivi	Source	Population	Date
Suivi actif			
Questionnaires annuels	Volontaires	Tous les volontaires	Depuis 1989
Examen de santé	CES	Tous	1999 - 2006
Banque matériel biologique	CES	Tous	1999 - 2006
Suivi passif			
Absentéisme avec causes médicales	SGMC	En activité	Depuis 1978
Cancers (registre)	SGMC	En activité	Depuis 1989
Cardio ischémiques (registre)	SGMC	En activité	Depuis 1989
Recours aux soins, consommation médicale	CMCAS	Tous	Depuis 2000
Données démographiques Historique de carrières	GPSO	En activité	Depuis 1989
Conditions de travail	SGMT	En activité	Depuis 1994
Matrice emploi exposition	SGMT	Tous	Carrière entière
Causes de décès	Inserm	Tous	Depuis 1989

Taux de réponse à l'autoquestionnaire annuel

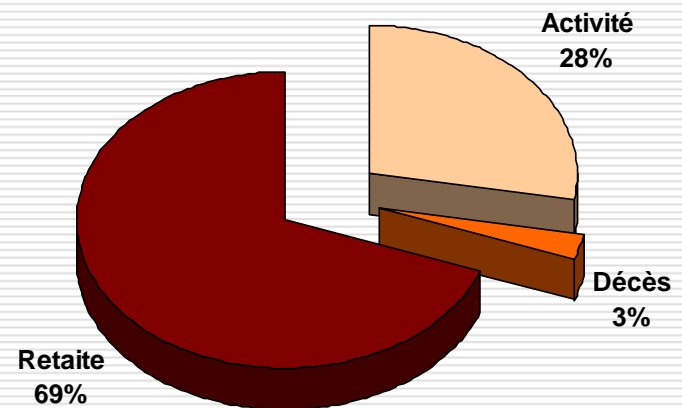


État de la cohorte GAZEL (31 décembre 2006)

HOMMES (15011)



FEMMES (5624)

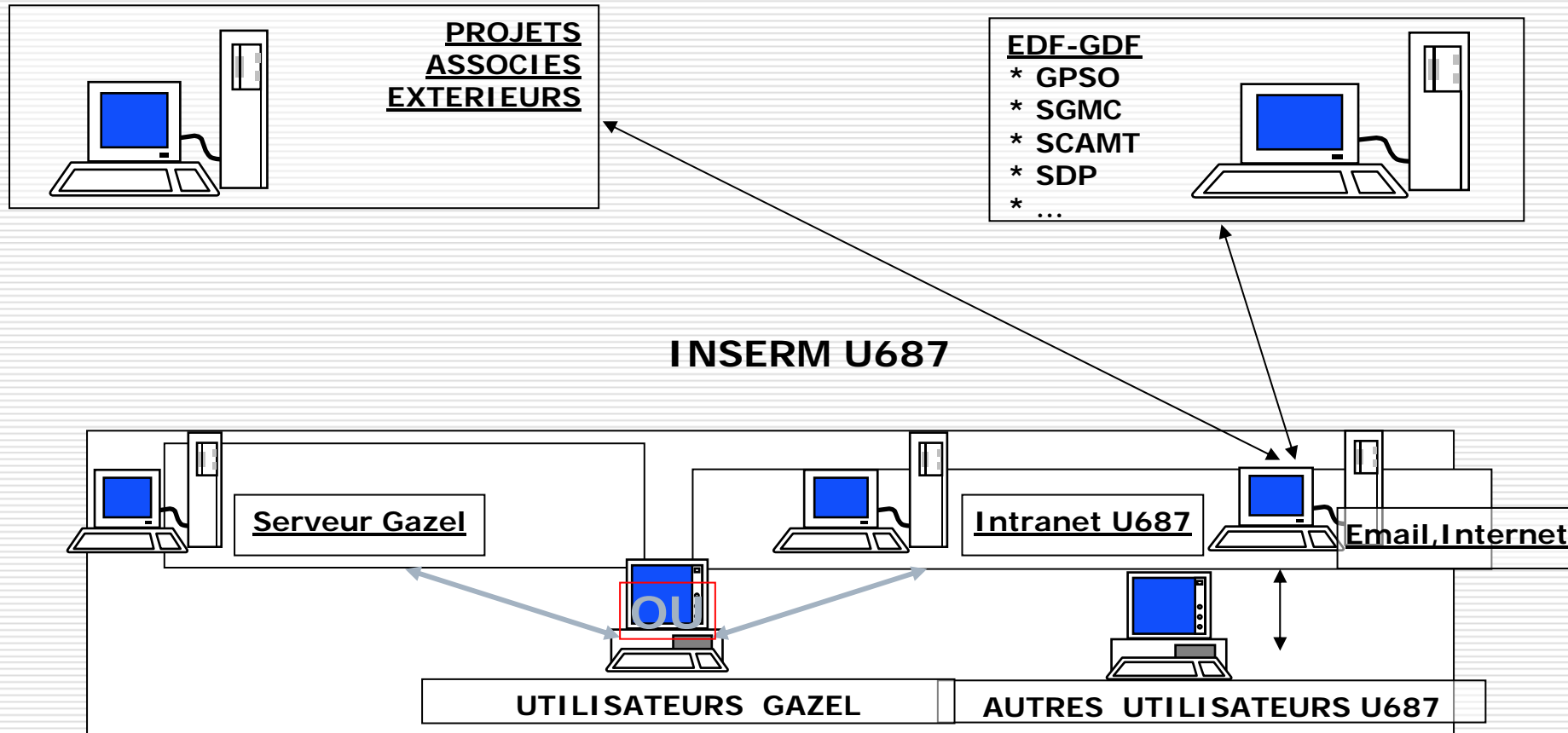


- 212 (1.0 %) ont demandé à ne plus recevoir de questionnaires
- 201 (0.9%) perdus de vue : 193 ont quitté EDF-GDF ; 8 demandes de suppression des fichiers



Confidentialité, secret médical et déontologie (1)

- ❑ La technique : séparation physique des données et de l'identification



Confidentialité, secret médical et déontologie (2)

- ❑ La loi et la Déontologie : accord de la Commission National Informatique et Liberté, du Conseil National de l'Ordre des Médecins, du Conseil National d'Éthique
- ❑ « Groupe de Suivi » formé par des représentants des Fédérations Syndicales représentatives du personnel EDF GDF
- ❑ Journal de la Cohorte (2 numéros/an), site Internet

La recherche épidémiologique dans GAZEL

□ Principes

- Tout chercheur peut accéder aux données et/ou aux sujets et bénéficier du suivi permanent de la cohorte
- Possibilité de recueil de données additionnelles
- Accord Cnil (et CPP le cas échéant) et du Comité scientifique de GAZEL

□ Actuellement : plus de 40 projets de recherche français et étrangers terminés ou en cours

Principaux thèmes de recherche

❑ **Problèmes de santé**

- Maladies chroniques (cardio-vasculaire, cancer,)
- Dépression et troubles mentaux communs diabète
- Ménopause et ostéoporose
- Migraine
- Troubles musculo-squelettiques
- Vieillesse cognitive et fonctionnelle
- Accidents de circulation
- Morbidité générale, absentéisme, mortalité

❑ **Facteurs de risques personnels**

- Tabac et alcool, alimentation
- Facteurs psychosociaux
- Événements de vie, support social
- Téléphone portable

❑ **Facteurs professionnels et sociaux**

- Expositions professionnelles
- Organisation du travail
- Pollution
- Répartition géographique des maladies
- Déterminants des inégalités sociales de santé

Équipes de recherche associées en France

- Cermes (Inserm, CNRS, EHSS), Villejuif
- Clinique du sommeil (CNRS), Bordeaux
- Inserm E361, Montpellier
- Inserm Unité 149, Villejuif
- Inserm Unité 435, Rennes
- Inserm Unité 558, Toulouse
- Inserm Unité 593, Bordeaux
- Inserm Unité 687, Saint-Maurice
- Inserm Unité 754, Villejuif
- Inserm Unité 780, Villejuif
- Institut des sciences de l'évolution, Montpellier
- Inrets, Lyon
- Ined, Paris
- Laboratoire Espace Santé et Territoire, Paris
- ORS Ile de France, Paris
- Service de Psychiatrie, Hôpital Européen Georges Pompidou, Paris
- Service Général de Médecine de Contrôle ; EDF-GDF, Paris
- Institut de veille sanitaire, Département Santé Environnement, Saint Maurice

Équipes de recherche étrangères associées

- Vrije Universiteit, Bruxelles, Belgique
- Catholic University of Louvain, Bruxelles, Belgique
- Department of Epidemiology, McGill University, Montreal, Canada
- Medical Faculty, University of Düsseldorf, Allemagne
- Department of Epidemiology and Public Health, London University College, Angleterre
- Harvard School of Public Health, Boston, USA
- National Institute of Psychosocial Medicine, Stockholm, Suède
- Karolinska Institute, Stockholm, Suède
- Finnish Institute of Occupational Health, Helsinki, Finlande
- School of Medicine, Yale, New Haven, USA
- School of Medicine, UCLA, Los Angeles, USA

Constances

Une nouvelle cohorte
longitudinale en préparation

Objectifs : une cohorte pour la santé publique et la recherche

- ❑ Santé publique : décrire
 - La santé de la population du régime général de sécurité sociale (80 % de la population)
 - Les parcours de prévention et de soins
 - Identifier des groupes à risque
 - Suivre certains indicateurs de la loi de santé publique
- ❑ Recherche : principalement autour de 4 thématiques
 - Risques professionnels (partenariat DST-InVS)
 - Vieillesse
 - Inégalités sociales de santé
 - Santé des femmes
- ❑ Ouverture à la communauté scientifique

Structure sociodémographique de la cohorte Constances

- ❑ Échantillon représentatif de structure proportionnelle à celle de la population du Régime général pour âge, sexe, PCS
- ❑ Adultes entre 18 et 70 ans
- ❑ Effectif : 200 000 personnes
- ❑ Inclusion graduelle sur 5 ans

Contexte

□ L'Assurance Maladie

- Le réseau des Centres d'Examens de Santé
- Les grandes Bases de données nationales de l'AM
 - Consommations de soin
 - ALD, AT-MP
 - PMSI (intégrée au SNIR-AM)

□ Les bases de données de la Cnav

□ Inserm : causes de décès (Cepi-DC)

Principe général : s'appuyer sur des structures existantes pour optimiser sur le plan des moyens et de la qualité des données

Le réseau des Centres d'Examens de Santé



- 110 Centres
- 650 000 bilans par an
- Centre Technique d'Appui et de Formation des CES (Cetaf)

Principes généraux du protocole

Inclusion au CES

- ❑ Auto questionnaires : données sociodémographiques, santé, modes de vie, vie professionnelle, sociale et familiale, expositions professionnelles
- ❑ Examen clinique, paraclinique et biologique standardisé
- ❑ Examen spécialisé pour les plus âgés (55 ans +)
 - Tests fonctionnels (tonus, marche, équilibre, etc.)
 - Tests cognitifs
- ❑ Banque de matériel biologique et génétique

Suivi

Actif

Par auto questionnaire annuel

- suivi serré des paramètres personnels
- recueil de nombreuses données sans surcharger les questionnaires
- réactivité par rapport à des questions de recherche

Problème des adresses : mises à jour dans différents fichiers (DADS, CAF, RNIAM), suivi par La Poste, fiche de suivi dans les autoquestionnaires

Examen CES : périodicité variable

Passif par les grandes bases de données nationales

SNIIR-AM : santé

Cnav : événements socioprofessionnels

CepiDc - Inserm : causes de décès

L'accès aux bases de données nationales

□ Les grands principes

- Accès au fichier de la CNAV par le NIR bénéficiaire
- Accès au SNIIR-AM par NIR ayant droit après :
 - FOIN 1 réversible,
 - FOIN 2, irréversible, détenu par la CNAMTS

□ Un tiers de confiance

- Gère la correspondance entre les fichiers nominatifs+NIR et les numéros d'anonymat
- Détient FOIN 1
- Algorithme de cryptage pour adresser les NIR à la Cnav

□ Pas de liaison entre les bases de données

Aspects méthodologiques

Représentativité et effets de sélection

Les 17 CES participant à Constances



Les Centres d'examen de santé de CONSTANCES

250 000 consultants / an

Représentativité géographique

La structure de la population des départements où sont situés les CES *CONSTANCES* est pratiquement identique à celle de la France entière pour les principales caractéristiques démographiques et socioprofessionnelles :

- Âge
- Sexe
- PCS
- Secteurs d'activité économique

Volontariat et effets de sélection

- Volontariat : effets de sélection inévitables à toutes les étapes
 - dans les études transversales (même si tirage au sort)
 - à l'inclusion et lors du suivi dans les enquêtes longitudinales

- Déterminants du volontariat
 - inclusion : importance de la situation sociale, de l'état de santé et des comportements à risque
 - attrition au long du suivi : rôle majeur des événements de santé

Mais volontariat indispensable pour le recueil de
certains types de données

Effets de sélection : les problèmes de biais

L'importance des biais est fonction de l'objectif

- ❑ Plus important dans un objectif d'analyse descriptive de l'état de santé : estimation de l'incidence ou la prévalence d'une maladie ou d'une exposition
- ❑ Moins important dans un objectif d'analyse étiologique (relation entre exposition/intervention et maladie) : pas de biais si les mécanismes de sélection sont similaires à l'inclusion et durant le suivi dans les groupes comparés

Effet de sélection : les solutions

Redressement par des méthodes de pondération

- ❑ Nécessite des informations pertinentes sur les variables connues pour être liées au volontariat :
 - sociodémographiques
 - santé
- ❑ Disponibles sur l'ensemble de la population cible et dans Constances

Construction de l'échantillon

- Plan de sondage stratifié avec probabilités inégales
 - Strates : 17 CES participants
 - Bases de sondage : Cnav pour les affiliés des CPAM concernées
 - Invitations après tirage au sort
 - Probabilité inégales : augmenter la probabilité de réponse de certaines catégories de personnes dont on sait qu'elles participent moins

- Redressement puis calage sur marges

Une cohorte « témoin »

- « Cohorte témoin » : échantillon tiré au sort dans la population de non répondants (effectif double de Constances)
 - Suivi dans les fichiers de la Cnav (données socio professionnelles)
 - Suivi dans le SNIIR-AM
 - hospitalisations
 - consommation détaillée de soins
 - ALD (fréquence, diagnostic)
- Possibilité de redressement et de calage sur marges au fur et à mesure du vieillissement de Constances

Calendrier prévisionnel

- Pilotes : 2008
- Inclusion : 2009