

---

# Méthodes de microsimulation: théorie et applications

---

Les Rencontres de statistique appliquée

Ined

31 mai 2018

Sophie Pennec

Institut national d'études démographiques (INED), F-75020 Paris, France



- Méthodes de microsimulation
- Exemple d'implémentation et de données mobilisées pour un modèle de microsimulation de dépendance des personnes âgées (DynoptaSim)

# Modélisation/projections

---

- Trois principaux types de modèles
  - Modèles analytiques
    - population stable; évolution par taux de croissance
  - Modèles de macro-simulation
    - modèles par composantes:  
$$P(t) = P(t-1) + (N-D) + (I-E)$$

N=naissances (tx fécondité);  
D=décès (quotient mortalité);  
I-E=solde migratoire
  - Modèles de micro-simulation

# Microsimulation: origine

---

- Guy Orcutt (physicien et économiste) définit les principes de la microsimulation à la fin des années 1950 (article fondateur en 1957; ouvrage avec première application en 1961)
- Il part du constat que les modèles analytiques et le niveau macro:
  - Ne permettent pas d'obtenir de bonnes estimations des effets de changements de politiques sociales ou fiscales (des gagnants et perdants car politiques non cohérentes quelque fois)
  - Ne peuvent pas bien prendre en compte les effets de variables qui sont liés, sur le comportement des individus
  - Ne peuvent pas reproduire correctement les différentiels (économiques, sociaux) et ne peuvent pas donner des estimations fines par groupe très détaillée de population

# Microsimulation: approche

---

- Prend une perspective différente
- La modélisation analytique donne des résultats pour « l'homme moyen » pour fournir des agrégats/résultats au niveau macro
- Microsimulation simule le comportement de chaque individu et c'est la somme des comportements individuels qui donne le résultat au niveau macro ou tout autre niveau d'agrégation souhaité

# Microsimulation : types de modèles

---

- Caractéristiques définissant les différents types de modèles de microsimulation
  - Temps
    - Modèles Statique vs. Dynamique
      - + Dynamique -> utilisation de processus stochastique – méthode de Monte Carlo
    - Modèles Discrets vs. Continus
  - Comportement des individus
    - Aucun
    - Déterminé par des probabilités
    - Déterminé par un marché
    - Stochastique ou déterministe
  - Modèle ouvert vs. Fermé
  - Modèle transversal ou longitudinal

# Microsimulation : Avantages

---

## Avantages de la microsimulation

- Donne des résultats non pas uniquement en termes de moyenne mais donne aussi la variance (forme d'intervalle de confiance)
- Crée des biographies
- Peut être utilisée comme outil d'expérimentations (what if)
- Peut prendre en compte de nombreuses variables et de nombreux événements
- Modèles construits en modules pour faciliter l'ajout de nouvelles covariables sans avoir à reconstruire complètement le modèle (modules ou sous-modules par événements/groupe d'événements)
- Tabulations des résultats très flexible: on peut avoir des résultats sur des sous-populations (niveau agrégation pas déterminé à la simulation)

# Microsimulation : Inconvénients

---

## Inconvénients de la méthode:

- Long à créer/mettre au point (pour les modèles de politiques sociales – plusieurs années-hommes)
- Besoin important de données: données individuelles, pour population initiale, pour estimer les transitions
- Pour gérer les événements concurrents (competing events)
- Variabilité des résultats (en raison des événements qui sont simulés de manière stochastique)
- Validation du modèle toujours longue



# Microsimulation : types d'utilisation

---

## → Modèles de politique sociale

- Retraite/dépendance: Destinie (France); SAGE (UK); SFB3 (Allemagne);
- Politiques sociales: APPSIM (Australia); Mosart (Norvège); LifePath (Canada)
- Modèle de santé: POHEM et ses successeurs (Canada); Sage; Pacsim

## → Modèle démographiques

- Les premiers : Simul (FR); DM1 – DM4 (Suède)
- Camsim (UK); Socsim (US);
- Famsim (Autriche);
- PopSim (Canada): projections officielles population par groupe linguistique

---

## DYNOPTA-SIM

Un modèle de microsimulation dynamique de la dépendance

---

Heather Booth, Binod Nepal, Sophie Pennec, Barbara Edgar,  
Richard Burns, Kim Kiely, Kaarin Anstey, Laurie Brown

Australian National University

# Sommaire

---

- Données DYNOPTA et modèle DYNOPTA-SIM
- Population initiale (création, validation)
- Modèle et Transitions
- Validation
- Résultats

- Projet d'une durée de 5 ans regroupant plusieurs universités australiennes
- Création de données harmonisées d'enquêtes longitudinales pour analyser différents facteurs de risques liés à la dépendance et pour créer un modèle de simulation de la dépendance

- Modéliser la dépendance des personnes de 45 ans et plus vivant en ménages ordinaires
- Simuler les facteurs de risques et les limitations en fonction de facteurs socio-démographiques, caractéristiques biomédicales
- Applications
  - Limitations multiples/durée avec limitations/en situation de dépendance
  - Espérance de vie sans ou avec incapacités
  - Demande d'aides (formelle ou informelle) à domicile

# Modèle

---

- Dynamique – modélise des cohortes de personnes âgées de 45+ en 1996 Individus (pas les ménages)
- Ouvert – uniquement les individus et ils ne sont pas liés entre eux ni regroupés en ménages
- Simulation jusqu'à extinction – sortie par l'entrée en institution ou décès
- Pas d'entrées – pas de migrations
- Modèle à temps discret – pas mensuel
- Transitions – processus de Markov de premier ordre

DYNOPTA-SIM

# CONSTRUCTION DE LA POPULATION INITIALE

# La base de données DYNOPTA

---

9 enquêtes longitudinales australiennes sur le vieillissement:

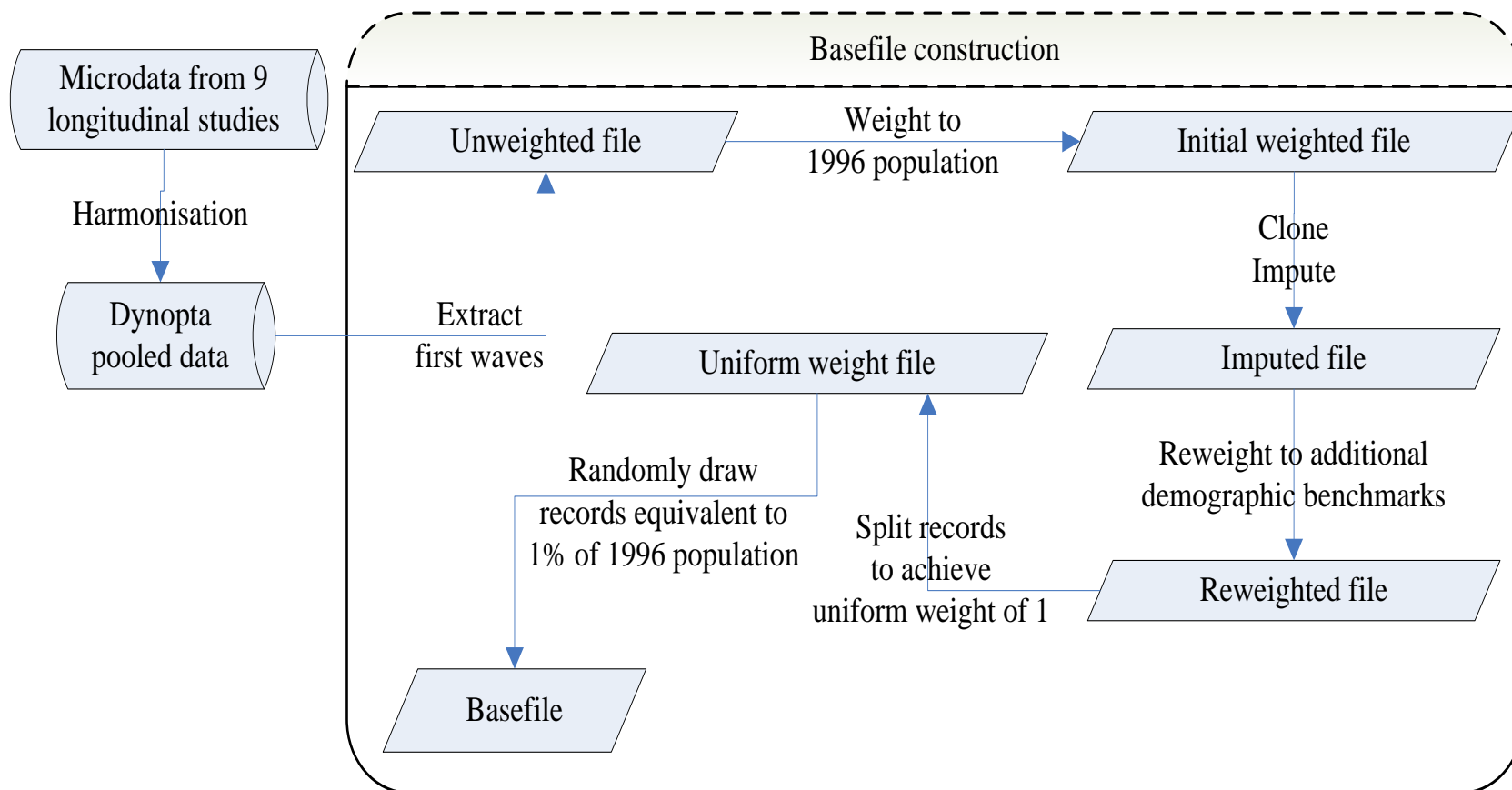
- 50 652 personnes âgées 45+ à la 1ère vague (1990-2002)
- Pondérations – age, sexe, région – cumul des premières vagues est représentatif de la population de 1996
- Les personnes en institution sont mises à part (3416 personnes)
- 47,236 personnes en ménages ordinaires
- Poids ajustés pour la population en ménages ordinaires



# La base de données DYNOPTA

Enquête	Années	Ages	N
Australian Longitudinal Study of Ageing (ALSA)	1992-93	64-103	1 961
Australian Longitudinal Study of Women's Health Mid Cohort (ALSWH-mid)	1996	45-51	12 653
Old Cohort (ALSWH-old)	1996	68-76	11 113
Australian Diabetes and Obesity and Lifestyle Study (AusDiab)	1999-2000	45-91	6 780
Blue Mountains Eye Study (BMES)	1992-93	45-100	3 508
Canberra Longitudinal Study (CLS)	1990-91	70-100	899
Household, Income and Labour Dynamics of Australia (HILDA)	2001-02	45-92	6 162
Melbourne Longitudinal Study Healthy Ageing (MELSHA)	1994	66-95	991
Personality and Total Health Through Life (PATH)	2001-02	60-66	2 550
Sydney Older Person's Study (SOPS)	1991-93	75-97	619
Total premières vagues empilées de DYNOPTA (non pondéré)	1990-2002	45-103	47 236

# Population initiale



Nom de la variable	Nombre d'enquêtes	Nombre valeurs manquantes	% manquants (N=47236)	Catégories des variables
<b>Socio-demographic</b>				
Age	10	0	0.0	Âge par année
Sex	10	0	0.0	Homme Femme
Union	10	53	0.1	Marié/de facto Divorcé ou séparé Veuf ou veuve Jamais marié
Education	10	3968	8.4	Education terminée avant 15 ans Education terminée à 15 ans ou plus
Profession	10	7708	16.3	Manager/professional Clerical/para-professional Tradesperson Sales/service/labourer
<b>attitude</b>				
Sociabilité (contact)	3	43808	92.7	Au moins une fois par semaine Moins d'une fois par semaine
Activité physique	5	33600	71.1	Aucune Moins d'une fois par semaine Une fois ou plus par semaine Chaque jour
Tabac	10	1714	3.6	Jamais Ancien fumeur Oui
Alcool	9	3805	8.1	Pas de risque lié à l'alcool Risque faible lié à l'alcool Risque fort et très fort lié à l'alcool

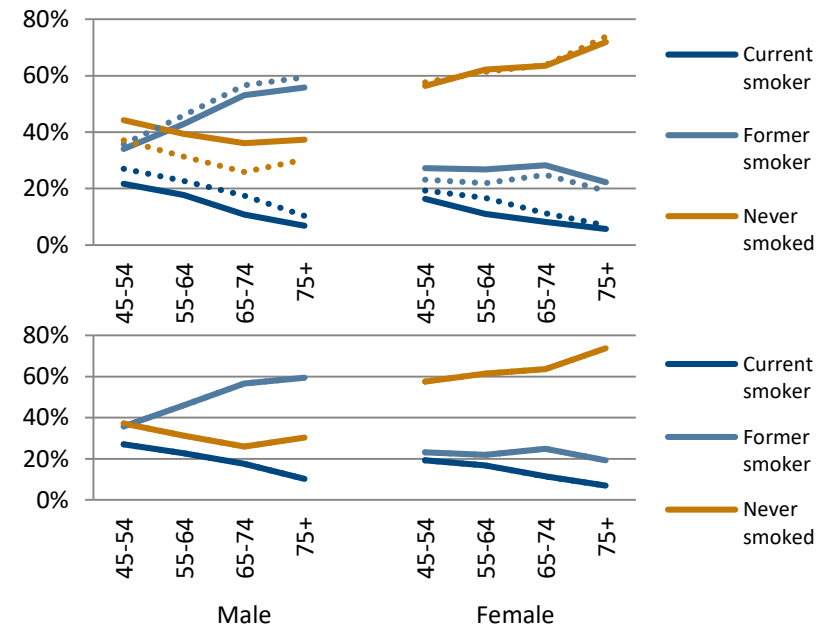
Nom de la variable	Nombre d'enquêtes	Nombre valeurs manquantes	% manquants (N=47236)	Catégories des variables
<b>Biomédicales</b>				
IMC	7	10762	22.8	Souspoids Normal Surpoids Obésité
Hypertension	9	6556	13.9	Jamais Oui
Diabète	9	6525	13.8	Jamais Oui
Stroke	9	6580	13.9	Jamais Oui

<b>Incapacités</b>				
Problèmes cognitifs	4	41317	87.5	Pas de limitation cognitive limites cognitives probables (MMSE<24)
Depression	9	4456	9.4	Non Dépression Probable
Problème de vue	5	38521	81.6	Non Oui
Problème d'audition	1	45735	96.8	Non Oui
Limitations physique	6	8548	18.1	Aucune limitation Limitation légère Limitation élevée
Chute	5	29553	62.6	Non Oui

# Population initiale – validation

- Sources de donnée exogènes
  - 1996 recensement
  - 1995 Enquête Santé
  - Enquêtes sur force de travail
  - 1998 Enquête sur incapacité, vieillissement et aidants
- Variables re-pondérées: situation de couple (80+), education, Tabac, IMC

Tabac: avant et après la répondération



External Source: National Health Survey 1995

DYNOPTA-SIM

# TRANSITIONS

# Transitions: covariables

VARIABLE DEPENDENT	PREDICTEURS
Couple	Age, sexe, statut de couple
Contacts	Age, sexe, activité physique, depression
Activité Physique	Age, sexe, statut de couple, activité physique, AVC
Tabac	Age, sexe, statut de couple, age de fin d'études, niveau consommation d'alcool
Consommation d'alcool	Age, sexe, statut de couple, age de fin d'études, fumeur
Indice masse corporelle	Age, sexe, age fin d'études
Hypertension	Age, sexe, niveau consommation d'alcool , IMC, fumeur
Diabetes	Age, sexe, fumeur , IMC, hypertension
AVC	Age, sexe, smoking, niveau consommation d'alcool
Démence	Age, sexe, age de fin d'étude
Depression	Age, sexe, statut de couple, age de fin d'études, fumeur, niveau consommation d'alcool
Problème de vue	Age, sexe, diabetes
Problème d'audition	Age, sexe, Activité professionnelle
Limitation physique	Age, sexe
Chute	Age, sexe, activité physique, problème de vue
Institutionnalisation	Age, sexe, statut de couple, démence

# Estimation de transition

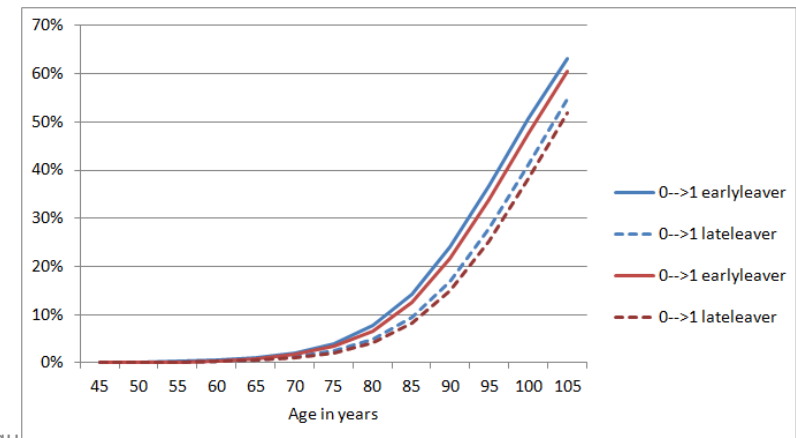
- vagues 1 et 2 (ou 3) des données de DYNOPTA (niveau ménage)
- vague 1 1990-2002
- vague 2 1994-2006
- Intervalle moyen = 32 mois
- Intervalle spécifique à chaque variable/transition

## Alzheimer/démence

	0 to 1
Femme	-0.157215
Age	0.144405
Age fin d'études > 15 ans	-0.498773
Intercept	-12.56220
Interval	39.85

Note: 0 no dementia, 1 probable dementia

## Probabilités de transition (Homme et Femme)





DYNOPTA-SIM

# VALIDATION

# Validation: données et comparaisons

---

- Données externes pour 2001 à 2016:
  - Recensements 2001; 2006; 2011
  - Enquêtes Santé 2007-8; 2014
- Taux de prévalence transversal – profil et évolution
- Taux de prévalence longitudinal – profil et évolution

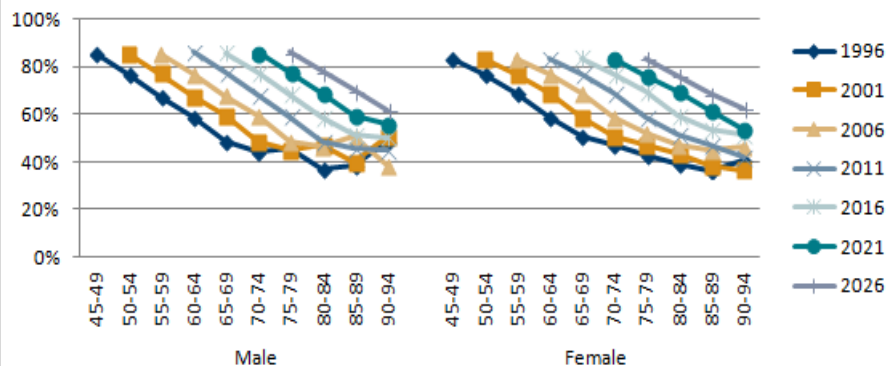
# Resultats provisoires – Démence

Effect of risk factors on Probable Dementia transition:

Covariables	0 to 1
Age	↑
Femme	↓
Education jusqu'à 15+	↓

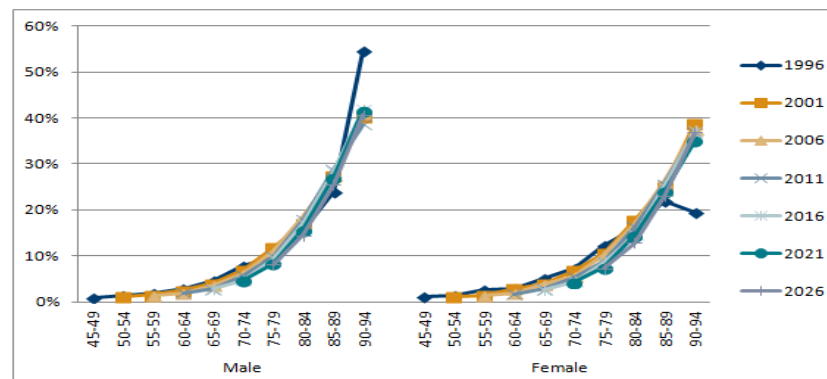
Left school age 15+

Résultats transversaux

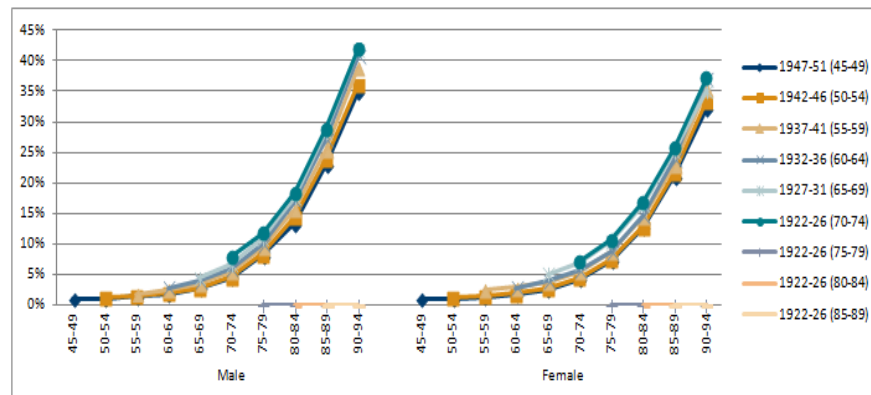


Probable Dementia

Résultats transversaux

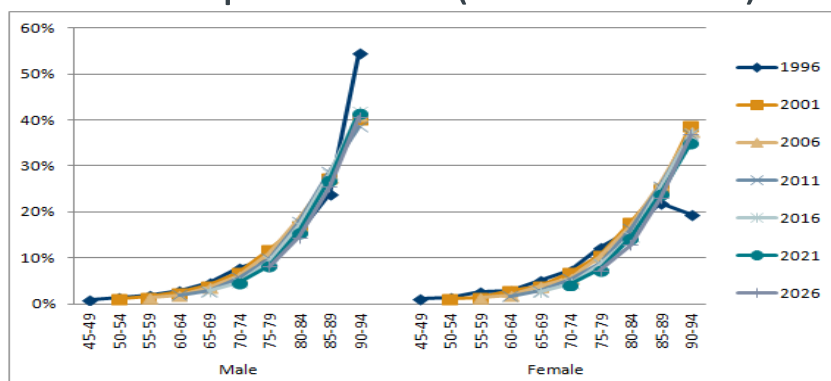


Résultats longitudinaux

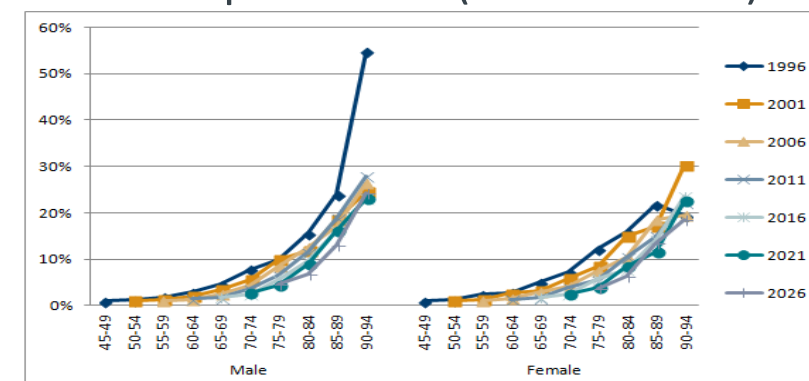


# Resultats – Démence

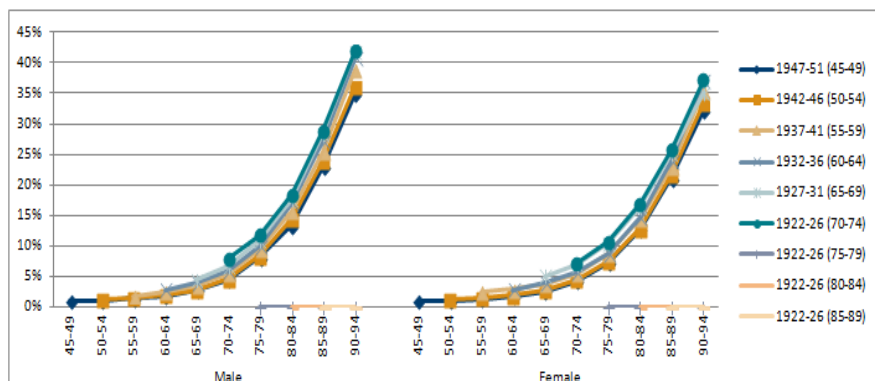
Avec les probabilités actuelles  
Résultats par année (transversaux)



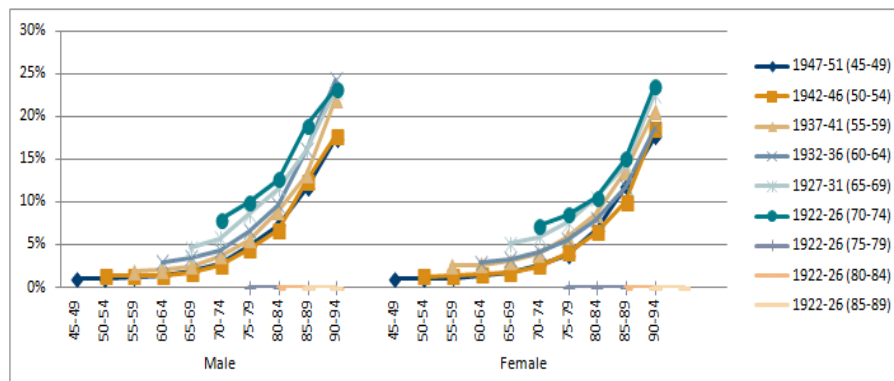
probabilités de 5 ans plus jeunes  
Résultats par année (transversaux)



Résultats par cohortes (longitudinaux)



Résultats par cohortes (longitudinaux)



# Etapes à venir

---

- Finir les validations
- Analyses
  - Effect d'interventions (scenarios “et si”)
  - Besoin d'aides formelles
- ....

Merci