



Sciences Economiques & Sociales de la Santé  
& Traitement de l'Information Médicale

<http://sesstim.univ-amu.fr>

***Margarita MORENO-BETANCUR***

*Victorian Centre for Biostatistics (ViCBiostat), Melbourne -AUSTRALIE*

*Centre d'épidémiologie sur les causes médicales de décès (CépiDc) - Inserm, Paris - FRANCE*

## **Définition et estimation de deux indicateurs synthétiques d'inégalités socioéconomiques de santé**

novembre 2016



**Cliquez ici pour voir l'intégralité des ressources associées à ce document**

# Définition et estimation de deux indices synthétiques d'inégalités socioéconomiques de santé

Margarita Moreno-Betancur

Moreno-Betancur M, Latouche A, Menvielle G, Kunst AE, Rey G.

**Relative index of inequality and slope index of inequality:**

**A structured regression framework for estimation.**

*Epidemiology* 2015; 26(4):518-527.

Webinar QuanTIM - Université Aix-Marseille

18 Novembre 2016

# Contexte

La réduction des inégalités de santé à l'intérieur des pays et entre les pays est l'une des priorités de l'Organisation Mondiale de la Santé (OMS)

ex. programmes "Health for all by the year 2000" et "Closing the Gap in 2008"

**Pour surveiller l'évolution de ces inégalités, il faut d'abord les définir et trouver un moyen de les mesurer**

# Définitions

- “Statut socioéconomique”:
  - Mesures ordonnées: Niveau d'études, revenu,..
  - Mesures non-ordonnées: Catégorie socioprofessionnelle,...
- “Santé”:
  - Incidence de maladies
  - Mortalité (toute-cause et cause-spécifique)
  - Indice de masse corporelle
  - Accès aux soins
  - ...
- “Inégalités”
  - Gradient=association linéaire (statut socioéconomique  $\propto$  santé)
  - Variabilité globale
  - ...

# Définitions

- Le choix de mesures de statut socioéconomique et de santé dépend souvent de la disponibilité des données
  - Pour beaucoup de pays (ex. pays de l'OCDE\*) il existe des données en population générale de **mortalité par niveau d'études**, y compris pour la France
- Toutes les différentes définitions d'inégalité sont d'intérêt
  - Nous considérons des méthodes pour mesurer les **gradients**.
  - D'autres mesures sont appropriées pour d'autres définitions d'« inégalité » (ex. Indice de dissimilarité)

# Propriétés souhaitées d'une telle mesure

- Synthétique, d'interprétation facile pour les décideurs
- Reflète les expériences de toute la population, non seulement de ceux dans les catégories extrêmes de l'échelle socioéconomique
- Inégalités relatives vs. absolues: Les deux sont importantes
  - Un taux 50% plus élevé d'un problème de sante rare peut être considéré (ou pas) moins important en termes de sante publique qu'un taux 10% plus élevé d'un problème de sante très fréquent*
- Comparabilité de la mesure entre populations, temporellement et géographiquement

# Echantillon Démographique Permanent (EDP)

Echantillon représentatif de la population (1%) établi par l'Institut National de la Statistique et des Etudes Economiques (INSEE): n=535 008

- **Niveau d'études** disponible à partir des recensements de population (1968, 1975, 1982, 1990, 1999), en 4 catégories:
  - Aucun
  - Certificat d'études primaires
  - Diplôme professionnel
  - $\geq$  Bac

# Echantillon Démographique Permanent (EDP)

- **Statut vital et date de décès** disponibles du registre civil
- **Causes de décès** disponibles à partir d'un appariement avec le registre national de causes médicales de décès (Inserm - CépiDc)
  - Certificat de décès au format international (OMS)
  - Codage du certificat selon la Classification International des Maladies (CIM) établie par l'OMS (10ème révision)

## Question de recherche:

**Comment ont évolué les inégalités éducationnelles de mortalité toute-cause en France entre 1982 et 2006?**

Mesurer les inégalités sur 3 périodes de 7 ans:

- 1982-1988
- 1990-1996
- 1999-2005

Pour chaque période, nous étudions la mortalité des individus qui étaient vivants et âgés 30-84 ans en début de chaque période

# Comment ont évolué les inégalités éducationnelles de mortalité toute-cause en France entre 1982 et 2006?

<i>n</i> =	250 483	275 648	302 520
décès =	26 065	24 045	21 951

1982 → 1988      1990 → 1996      1999 → 2005

Recensements : recueil du niveau d'études

Nous voudrions comparer la mortalité des individus ayant différents niveaux d'études pour chaque période

# Comment modéliser la mortalité?

Les délais de survie  $T_i$ ,  $i = 1, \dots, n$ , sont sujets à des phénomènes de **censure** et **troncature**: nous ne les observons pas pour toutes les personnes

Nous faisons appel aux méthodes d'**analyse de survie**, qui servent lorsque l'on étudie des délais de survenue d'un événement (dans cet exemple, le décès)

Elles permettent d'estimer le risque de l'événement et le taux de l'événement en présence de ces phénomènes

# Comment modéliser la mortalité en fonction du niveau d'études?

Nous considérons des modèles de régression pour le « **hazard** », qui est le taux instantané de mortalité:

$$\lambda(t) = \lim_{dt \rightarrow 0} P(T \in [t, t + dt] | T \geq t) / dt$$

Un modèle classique est le modèle de Cox, qui permet d'obtenir des mesures relatives de l'effet du niveau d'études sur la mortalité

# Modèle de Cox

Soit

- $Z_0 = 1$  si « Aucun diplôme »,  $Z_0 = 0$  sinon
- $Z_1 = 1$  si « Certificat d'études primaires (CEP) »,  $Z_1 = 0$  sinon
- $Z_2 = 1$  si « Diplôme professionnel »,  $Z_2 = 0$  sinon
- $Z_3 = 1$  si «  $\geq$  Bac »,  $Z_3 = 0$  sinon

$$\lambda(t) = \lambda_0(t) \exp(\gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3)$$

## Interpretation

$\lambda_0(t)$  = hazard parmi ceux avec  $Z_0 = 1$  i.e. « Aucun diplôme »

$\exp(\gamma_1)$  = hazard ratio comparant « CEP » avec « Aucun diplôme »

$\exp(\gamma_2)$  = hazard ratio comparant « Dip. prof » avec « Aucun diplôme »

$\exp(\gamma_3)$  = hazard ratio comparant «  $\geq$  Bac » avec « Aucun diplôme »

# Propriétés souhaitées d'une telle mesure

- ✗ • Synthétique, d'interprétation facile pour les décideurs
- ✓ • Reflète les expériences de toute la population, non seulement de ceux dans les catégories extrêmes de l'échelle socioéconomique
- ✗ • Inégalités relatives vs. absolues: Les deux sont importantes
  - Un taux 50% plus élevé d'un problème de sante rare peut être considère (ou pas) moins important en termes de sante publique qu'un taux 10% plus élevé d'un problème de sante très fréquent*
- ✗ • Comparabilité de la mesure entre populations, temporellement et géographiquement

# Problème de comparabilité

Les différences dans la distribution du niveau d'études à travers le temps ou entre les populations changent la signification des catégories elles-mêmes à travers les populations

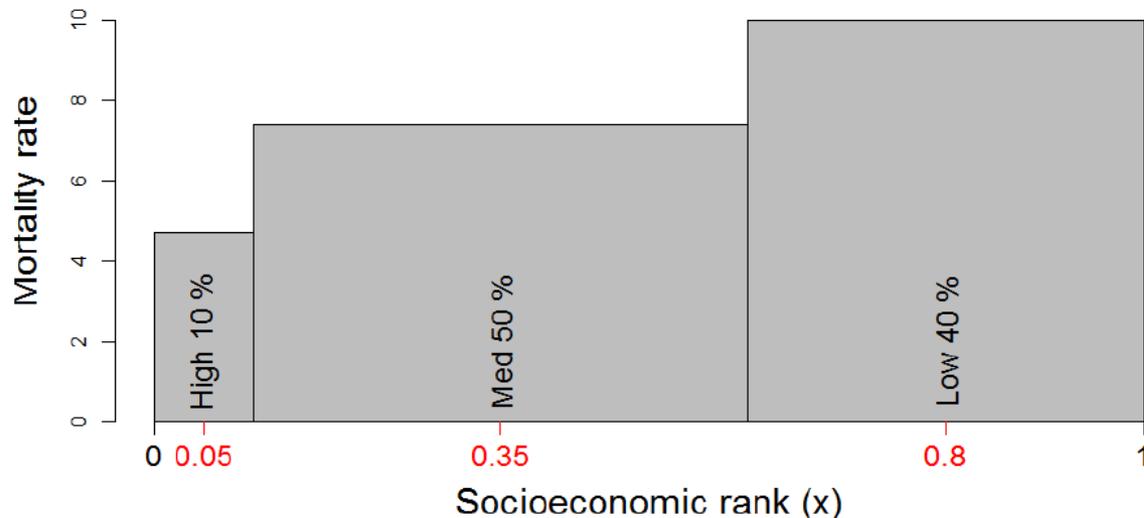
Ex: Etre sans diplôme aujourd'hui n'est pas du tout pareil que l'avoir été il y a 50 ans car ceci était plus commun auparavant, et par conséquent la promotion sociale parmi les gens sans diplôme était beaucoup plus fréquente

**Nous devons prendre en compte la position relative de l'individu dans l'échelle socioéconomique**

# Vers une mesure comparable

On définit le **rang socioéconomique** d'un individu comme le pourcentage de population qui est en-dessous de son niveau d'étude

En pratique, on doit approximer ce rang:



# Indices relatif et absolu d'inégalités (RII et SII pour les sigles en anglais)

Soit  $x$  le rang socioéconomique

Pour une mesure relative, on utilise un modèle de Cox:

$$\lambda(t) = \lambda_0(t) \exp(\beta x)$$

$$\text{RII:} = \exp(\beta^*)$$

Pour une mesure absolue, on utilise un modèle additif:

$$\lambda(t) = \lambda_0(t) + \alpha x$$

$$\text{SII:} = \alpha^*$$

# Indices relatif et absolu d'inégalités (RII et SII pour les sigles en anglais)

$\beta^*$ ,  $\alpha^*$  sont les paramètres « les moins mauvais »  
(*least false parameters*)

$$\text{RII} := \exp(\beta^*)$$

Gradient relatif moyen

$$\text{SII} := \alpha^*$$

Gradient absolu moyen

Ces nouvelles définitions correspondent à l'objectif d'obtenir des mesures synthétiques du gradient

# Propriétés souhaitées d'une telle mesure

- ✓ • Synthétique, d'interprétation facile pour les décideurs
- ✓ • Reflète les expériences de toute la population, non seulement de ceux dans les catégories extrêmes de l'échelle socioéconomique
- ✓ • Inégalités relatives vs. absolues: Les deux sont importantes

*Un taux 50% plus élevé d'un problème de sante rare peut être considère (ou pas) moins important en termes de sante publique qu'un taux 10% plus élevé d'un problème de sante très fréquent*
- ✓ • Comparabilité de la mesure entre populations, temporellement et géographiquement

# Méthodes d'estimation

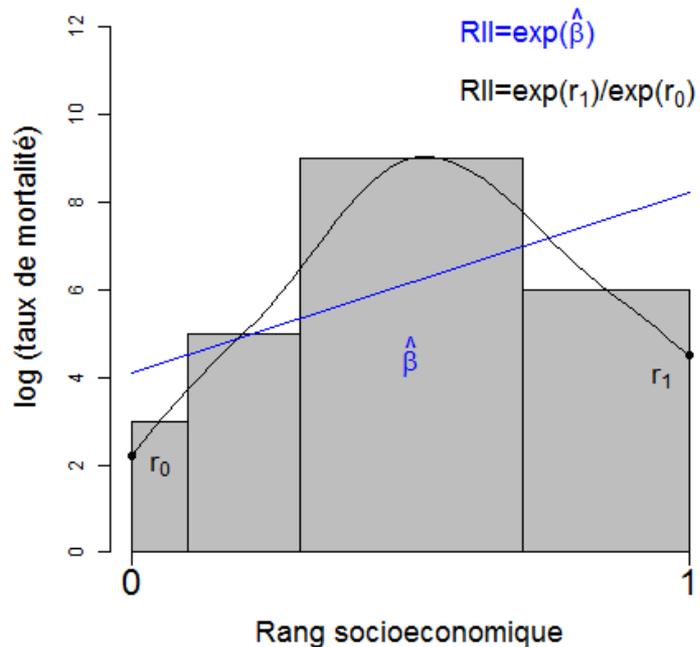
Nous avons proposé des méthodes d'estimation pour chaque indice avec différentes types de données (délais de survie individuels et données d'incidence agrégées)

TABLE 1. Regression Models for Estimation of the RII and SII With Event-rate Outcomes

Available Data	Health Outcome	Estimation	
		RII	SII
Individual time-to-event data	<i>Single event:</i> hazard rate	Cox model	Additive hazards model
	<i>Competing risks:</i> cause-specific hazard rate		
Event data aggregated by socioeconomic group	<i>Single event:</i> incidence rate	Multiplicative Poisson model	Additive Poisson model
	<i>Competing risks:</i> cause-specific incidence rate		

RII indicates relative index of inequality; SII, slope index of inequality.

# Comparaison avec l'usage d'avant de ces indices



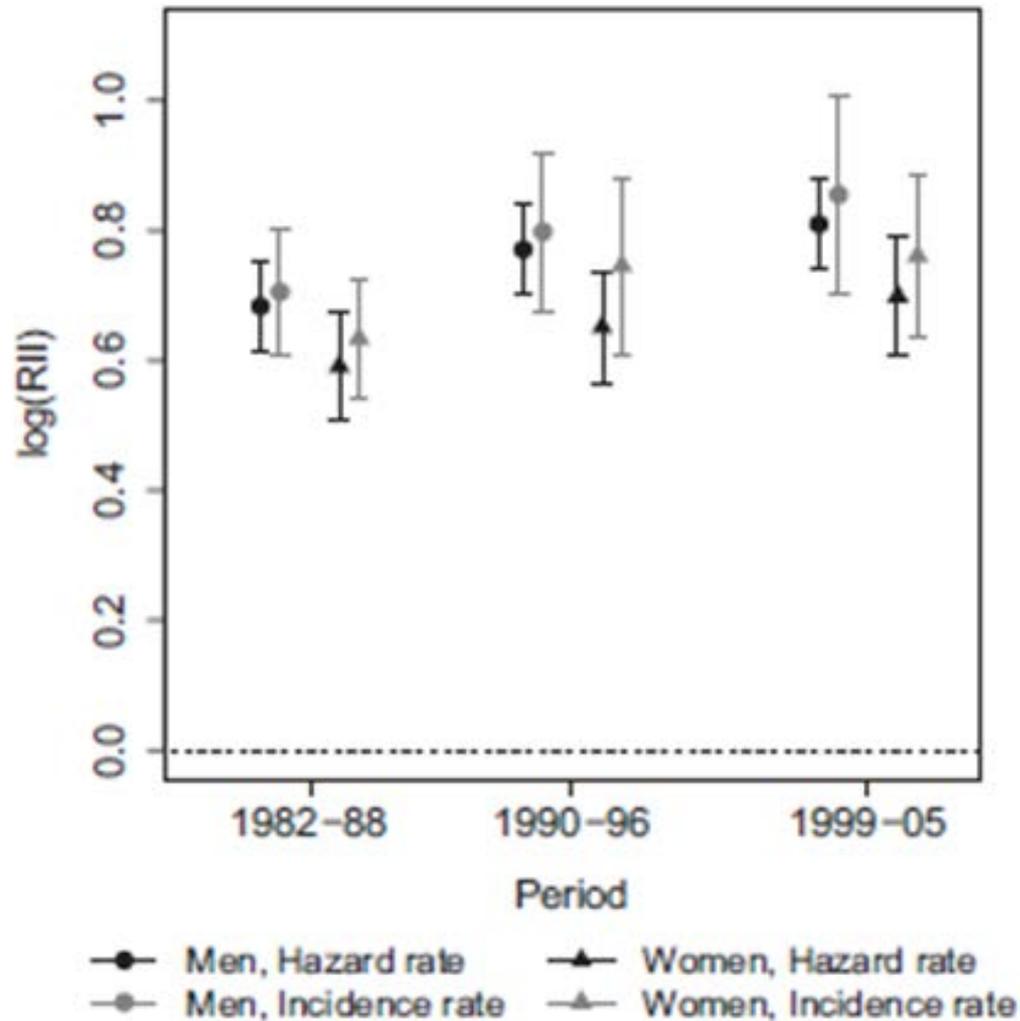
- RII était interprété à tort comme le risque relatif entre les scores 1 et 0
- SII était souvent calculé à l'aide d'une formule sans fondement théorique :

$$SII = \frac{2 \cdot \text{taux} \cdot (RII - 1)}{RII + 1}$$

Dans notre proposition il y a une claire distinction, mais aussi une harmonie, entre paramètre à estimer, données disponibles et méthode d'estimation

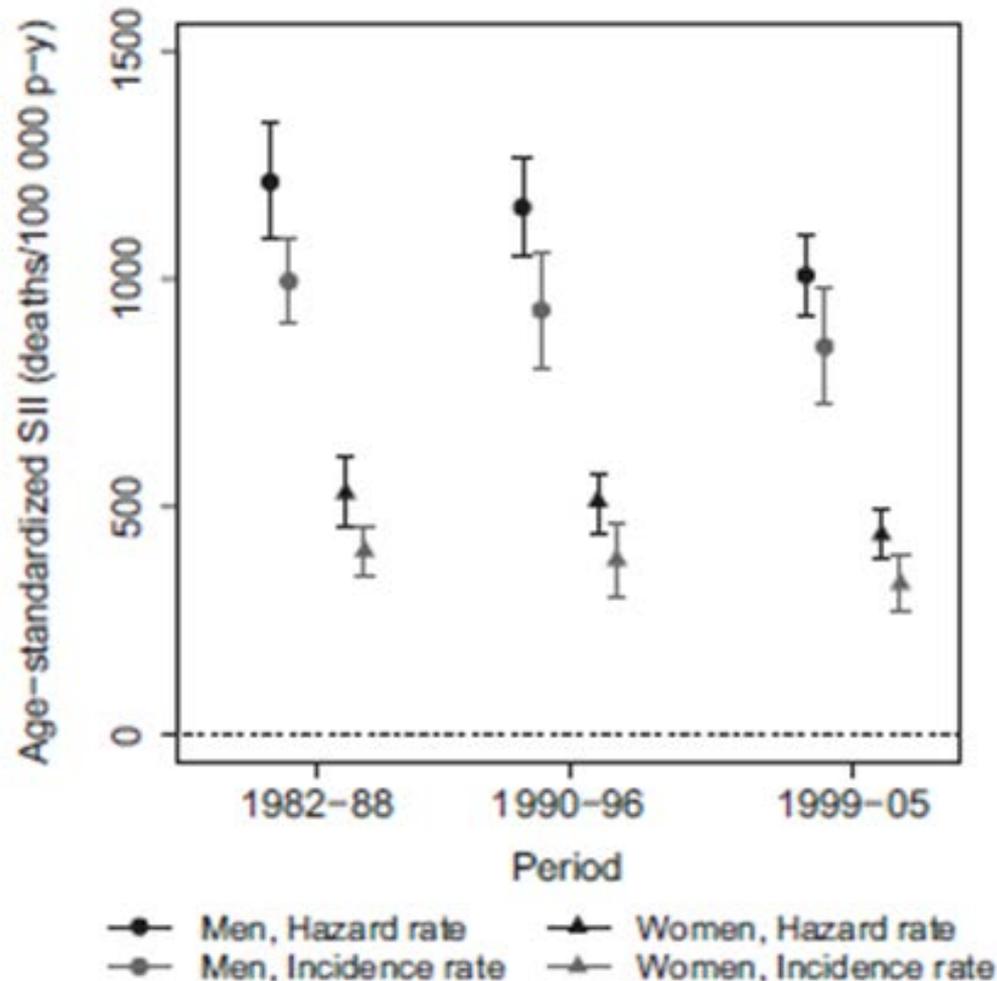
# Résultats pour l'EDP: mortalité toute-cause

Les inégalités relatives augmentent, et sont plus élevées pour les hommes



# Résultats pour l'EDP: mortalité toute-cause

Les inégalités absolues diminuent, et sont aussi plus élevées pour les hommes



# Mortalité cause-spécifique

Ex. Mortalité par cancer, par cause cardiovasculaire, etc.

Nous considérons des modèles de régression pour le **hazard cause-spécifique** :

$$\lambda_j(t) = \lim_{dt \rightarrow 0} P(T \in [t, t + dt], \mathbf{D} = \mathbf{j} | T \geq t) / dt$$

où  $\mathbf{D}$  est l'indicateur de la cause de décès

i.e.  $\mathbf{D} = \mathbf{j}$  indique la cause de décès  $j$  ( $j = 1, \dots, J$ )

# Mortalité cause-spécifique

Pour le RII, on utilise un modèle de Cox:

$$\lambda_j(t) = \lambda_{0j}(t) \exp(\beta_j x)$$

$$\text{RII}_j := \exp(\beta_j^*)$$

Pour le SII, on utilise un modèle additif:

$$\lambda_j(t) = \lambda_{0j}(t) + \alpha_j x$$

$$\text{SII}_j := \alpha_j^*$$

pour  $j = 1, \dots, J$

# Mortalité cause-spécifique

Pour le SII on a la propriété suivante

$$SII = \sum_{j=1, \dots, J} SII_j$$

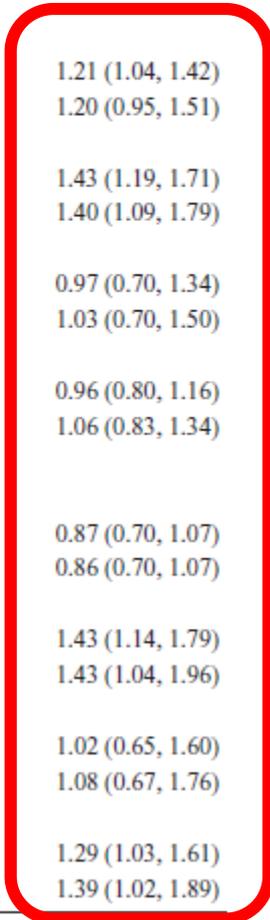
Donc la mesure suivante indique le pourcentage du SII toute-cause qui est attribuable à la cause  $j$ :

$$100 \times \frac{SII_j}{SII}$$

# Résultats pour l'EDP: mortalité cause-spécifique

TABLE 2. Estimates of the RII and 95% CI Obtained With Cause-specific Hazard and Incidence Rates for Each Period and Major Cause-of-death Groups

	1982–1988 <sup>a</sup>	1990–1996 <sup>a</sup>	1999–2005 <sup>a</sup>	Change (First to Last Period) <sup>b</sup>
	RII (CI)	RII (CI)	RII (CI)	RII Ratio (CI)
<b>Men</b>				
<b>Cancer</b>				
Hazard rate	1.76 (1.56, 1.98)	1.81 (1.62, 2.02)	1.91 (1.71, 2.14)	1.21 (1.04, 1.42)
Incidence rate	1.83 (1.59, 2.10)	1.82 (1.53, 2.17)	1.98 (1.67, 2.34)	1.20 (0.95, 1.51)
<b>Cardiovascular</b>				
Hazard rate	1.65 (1.47, 1.86)	1.99 (1.76, 2.26)	2.36 (2.04, 2.72)	1.43 (1.19, 1.71)
Incidence rate	1.71 (1.48, 1.98)	2.08 (1.71, 2.52)	2.40 (1.93, 2.98)	1.40 (1.09, 1.79)
<b>External</b>				
Hazard rate	2.40 (1.89, 3.05)	2.89 (2.30, 3.64)	2.50 (1.96, 3.19)	0.97 (0.70, 1.34)
Incidence rate	2.37 (1.85, 3.03)	2.97 (2.24, 3.92)	2.63 (1.84, 3.77)	1.03 (0.70, 1.50)
<b>Other</b>				
Hazard rate	2.71 (2.36, 3.11)	2.81 (2.43, 3.25)	2.62 (2.29, 3.01)	0.96 (0.80, 1.16)
Incidence rate	2.70 (2.29, 3.19)	2.97 (2.42, 3.63)	2.89 (2.38, 3.52)	1.06 (0.83, 1.34)
<b>Women</b>				
<b>Cancer</b>				
Hazard rate	1.16 (0.99, 1.37)	1.32 (1.13, 1.54)	1.04 (0.89, 1.20)	0.87 (0.70, 1.07)
Incidence rate	1.16 (0.99, 1.37)	1.41 (1.21, 1.65)	1.03 (0.89, 1.20)	0.86 (0.70, 1.07)
<b>Cardiovascular</b>				
Hazard rate	1.98 (1.73, 2.27)	2.27 (1.95, 2.65)	2.97 (2.46, 3.58)	1.43 (1.14, 1.79)
Incidence rate	2.15 (1.84, 2.52)	2.54 (2.07, 3.12)	3.39 (2.60, 4.42)	1.43 (1.04, 1.96)
<b>External</b>				
Hazard rate	1.83 (1.32, 2.54)	1.95 (1.41, 2.70)	2.33 (1.62, 3.36)	1.02 (0.65, 1.60)
Incidence rate	1.85 (1.31, 2.62)	2.14 (1.46, 3.15)	2.71 (1.84, 4.00)	1.08 (0.67, 1.76)
<b>Other</b>				
Hazard rate	2.40 (2.04, 2.82)	2.41 (2.03, 2.85)	3.07 (2.58, 3.65)	1.29 (1.03, 1.61)
Incidence rate	2.69 (2.13, 3.40)	2.92 (2.30, 3.72)	3.79 (3.05, 4.72)	1.39 (1.02, 1.89)



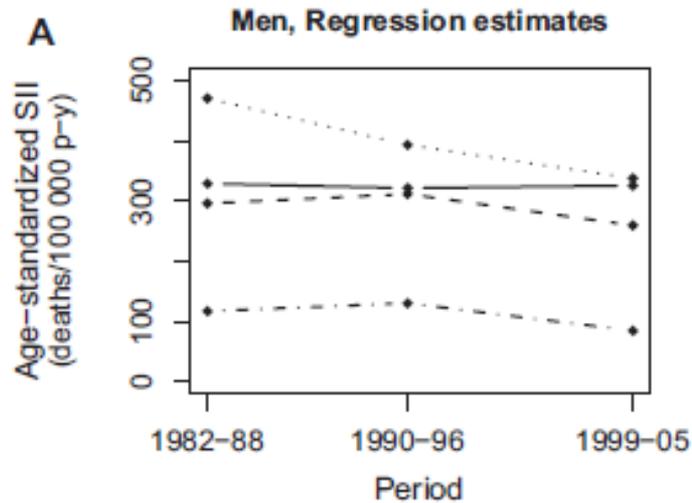
# Résultats pour l'EDP: mortalité cause-spécifique

TABLE 3. Estimates of the Age-standardized SII in Deaths Per 100,000 Person-years, 95% CI, and Percentage (%) Contribution to the All-cause SII, Obtained With Cause-specific Hazard and Incidence Rates for Each Period and Major Cause-of-death Groups

	1982–1988 <sup>a</sup>		1990–1996 <sup>a</sup>		1999–2005 <sup>a</sup>		Change (First to Last Period) <sup>b</sup>
	SII (CI)	%	SII (CI)	%	SII (CI)	%	SII Diff. (CI)
<b>Men</b>							
<b>Cancer</b>							
Hazard rate	329 (262, 404)	27	322 (254, 390)	28	326 (272, 381)	32	-3 (-87, 81)
Incidence rate	306 (250, 362)	31	279 (201, 356)	30	287 (236, 338)	34	-19 (-96, 59)
<b>Cardiovascular</b>							
Hazard rate	296 (228, 363)	24	312 (257, 366)	27	260 (212, 303)	26	-36 (-114, 41)
Incidence rate	231 (179, 284)	23	243 (203, 284)	26	189 (146, 233)	22	-42 (-116, 32)
<b>External</b>							
Hazard rate	117 (84, 147)	10	131 (104, 160)	11	85 (60, 108)	8	-32 (-74, 9)
Incidence rate	102 (73, 132)	10	113 (78, 148)	12	92 (58, 126)	11	-11 (-57, 36)
<b>Other</b>							
Hazard rate	470 (403, 539)	39	393 (340, 437)	34	338 (290, 379)	33	-132 (-206, -59)
Incidence rate	348 (301, 395)	35	286 (227, 345)	31	281 (233, 328)	33	-67 (-134, -1)
<b>Women</b>							
<b>Cancer</b>							
Hazard rate	37 (-1, 77)	7	78 (43, 118)	15	10 (-26, 43)	2	-28 (-83, 28)
Incidence rate	27 (-6, 61)	7	69 (35, 103)	18	15 (-14, 45)	5	-12 (-57, 33)
<b>Cardiovascular</b>							
Hazard rate	223 (184, 267)	42	196 (162, 231)	38	177 (150, 206)	40	-46 (-95, 3)
Incidence rate	169 (140, 197)	42	133 (107, 158)	35	111 (86, 135)	34	-58 (-96, -20)
<b>External</b>							
Hazard rate	38 (20, 57)	7	41 (22, 57)	8	34 (18, 47)	8	-4 (-27, 19)
Incidence rate	28 (11, 45)	7	35 (18, 52)	9	34 (21, 48)	11	6 (-15, 28)
<b>Other</b>							
Hazard rate	232 (194, 268)	44	199 (166, 234)	39	218 (185, 249)	50	-14 (-62, 34)
Incidence rate	181 (148, 214)	45	139 (101, 178)	37	162 (136, 187)	50	-19 (-61, 22)



# Résultats pour l'EDP: mortalité cause-spécifique



— Cancer      ···· External  
-- Cardiovascular      ···· Other

# Autres considérations dans l'estimation

- Ajustement par des facteurs de confusion
  - Standardisation par l'âge : nécessaire dans le cas du SII car les modèles additifs ne suffisent pas à ajuster la relation exponentielle entre l'âge et la mortalité
- Calcul du rang socioéconomique
  - Par strate, lors des analyses stratifiées (ex. par sexe)
  - Par groupe d'âge ou par cohorte de naissance, si l'on considère que c'est la position dans l'échelle sociale par rapport aux gens de la même génération qui est importante
- Extensions à d'autres types de mesure de santé ou métriques
  - Réponse binaire
  - Années de vie perdues par cause de décès

# “Take-home message”

Lorsqu'on choisit une méthode statistique pour répondre à une **question de recherche**, il faut faire une claire distinction entre le **paramètre à estimer, données disponibles** et la **méthode d'estimation**

Merci de votre attention !