

# Analyse de survie : Introduction

Olivier Bouaziz

`olivier.bouaziz@parisdescartes.fr`

`http://www.math-info.univ-paris5.fr/~obouaziz`

## Bibliographie

- ▶ Hill, C., Com-Nougué, C., & Kramar, A. (1990). Analyse statistique des données de survie. INSERM; Médecine-Sciences-Flammarion.
- ▶ Fleming, T. R., & Harrington, D. P. (1991). Counting processes and survival analysis. John Wiley & Sons.
- ▶ Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (1993). Statistical models based on counting processes. NY Springer.
- ▶ Kalbfleisch, J. D., & Prentice, R. L. (2002). The statistical analysis of failure time data. John Wiley & Sons.
- ▶ Collett, D. (2003). Modelling Survival Data in Medical Research. Chapman et Hall/CRC.
- ▶ Kleinbaum, D.G. & Klein, M. (2005). Survival analysis: a self-learning text. Springer Science & Business Media.
- ▶ Martinussen T., et Scheike, T. H. (2006). Dynamic Regression models for Survival Data. Springer Science & Business Media.

Types de données en analyse de survie - la  
censure à droite

# Données de suivi de patients

En médecine ou en biologie, on s'intéresse souvent à des durées :

1. Durée de survie de patients ayant eu un infarctus
2. Durée de rémission d'une leucémie aigüe
3. Durée de séropositivité sans symptôme de patients infectés par le VIH
4. Durée de fièvre chez un patient atteint de pneumonie

On distingue **l'évènement d'intérêt**

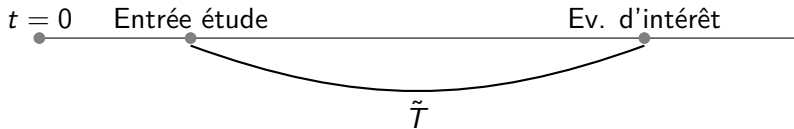
1. Décès du patient après l'infarctus
2. Fin de la rémission
3. Début des symptômes chez un patient séropositif
4. Fin de la fièvre chez un patient atteint de pneumonie,

de la **variable à expliquer** ou **variable d'intérêt** : durée avant l'apparition de l'évènement d'intérêt

1. Temps écoulé avant le décès
2. Temps écoulé avant la fin de la rémission
3. Temps écoulé sans symptôme
4. Temps écoulé avant la fin de la fièvre.

D'un point de vue statistique on note  $\tilde{T}$  la variable aléatoire d'intérêt.

- ▶ Elle représente un **temps**, c'est donc une variable aléatoire **continue** et **positive**.
- ▶  $\tilde{T}$  représente le temps écoulé depuis l'entrée dans l'étude et l'apparition de l'évènement d'intérêt.

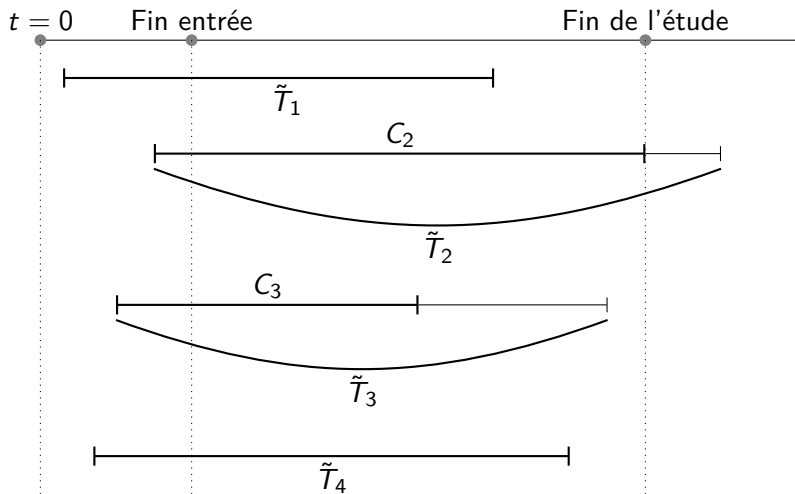


- ▶  $t = 0$  représente le début de l'étude.

## La censure à droite

- ▶ Dans une étude où l'on suit les patients au cours du temps (“follow-up study” en anglais), les patients rentrent à des instants différents et l'évènement d'intérêt se produit à des instants différents.
- ▶ Chez certains patients, on n'observe pas le moment où se produit l'évènement d'intérêt à cause de la **censure à droite** :
  - ▶ patient perdu de vue (il quitte l'étude spontanément, il déménage ...)
  - ▶ l'étude s'est terminée avant que l'évènement ne se produise (censure administrative).

## La censure à droite : un exemple sur quatre patients





## La censure à droite : un exemple sur quatre patients

Dans cet exemple,

- ▶ pour l'individu 1, la variable d'intérêt  $\tilde{T}_1$  est observée.
- ▶ pour l'individu 2, la variable d'intérêt  $\tilde{T}_2$  n'est pas observée. Seule la variable  $C_2$  est observée ! C'est un cas de **censure administrative**.
- ▶ pour l'individu 3, la variable d'intérêt  $\tilde{T}_3$  n'est pas observée (individu **perdu de vue**). Seule la variable  $C_3$  est observée !
- ▶ pour l'individu 4, la variable d'intérêt  $\tilde{T}_4$  est observée.

## La censure à droite

- ▶ On est en présence de censure à droite quand **l'évènement d'intérêt n'est pas toujours observé**.
- ▶ Pour les individus censurés, on observe une durée **plus petite** que la variable d'intérêt.

D'une manière générale, on note les observations, pour  $i = 1, \dots, n$ , de la façon suivante :

$$\begin{cases} T_i = \min(\tilde{T}_i, C_i) \\ \Delta_i = I(\tilde{T}_i \leq C_i). \end{cases}$$

## Exemple 1 : les données de Freireich (1963)

Durées de remission (en semaines) obtenue par des stéroïdes chez des patients atteints de leucémie aiguë, traités soit par placebo soit par 6-mercaptopurine (6-MP).

6-MP	6	6	6	6 <sup>+</sup>	7	9 <sup>+</sup>	10	10 <sup>+</sup>	11 <sup>+</sup>	13		
	16	17 <sup>+</sup>	19 <sup>+</sup>	20 <sup>+</sup>	22	23	25 <sup>+</sup>	32 <sup>+</sup>				
	32 <sup>+</sup>	34 <sup>+</sup>	35 <sup>+</sup>									
Placebo	1	1	2	2	3	4	4	5	5	8	8	8
	8	11	11	12	12	15	17	22	23			

Le signe + correspond à des patients qui ont quitté l'étude à la date considérée. Pour ces individus les durées sont donc **censurées à droite**. Par exemple le 4 ième patient est perdu de vue au bout de 6 semaines de traitement avec le 6-MP : il a donc une durée de rémission supérieure à 6 semaines.

## Exemple 2 : les données de Peto (1979)

Durées de survie de deux groupes de patients à qui l'on a administré deux types de traitement. Information supplémentaire : la fonction rénale, connue pour influencer sur la survie.

Durée observée	Groupe de traitement	Fonction rénale	Durée observée	Groupe de traitement	Fonction rénale
8	1	A	220	1	N
8	1	N	365+	1	N
13	2	A	632	2	N
18	2	A	700	2	N
23	2	A	852+	1	N
52	1	A	1296	2	N
63	1	A	1296+	1	N
63	1	A	1328+	1	N
70	2	N	1460+	1	N
76	2	N	1976+	1	N
180	2	N	1990+	2	N
195	2	N	2240+	2	N
210	2	N			

Les données avec + sont **censurées**. N: fonction rénale normale, A : fonction rénale anormale.

## Les données censurées requièrent un traitement particulier

Si on enlève les données censurées on perd de l'information !

- ▶ Dans l'exemple 1, si on enlève les données censurées, on ne tient pas compte des durées de rémission les plus longues et on sous évalue l'effet du traitement 6-MP.
- ▶ Dans l'exemple 2, si on enlève les données censurées, c'est à dire les 8 valeurs censurées on ne tient pas compte des patients qui ont justement les durées de survie les plus longues.

Si l'on ne prend pas en compte la censure, en faisant comme si une donnée censurée est égale à notre variable d'intérêt on aura tendance à sous évaluer les durées !

## Approche naïve : justification mathématique

- ▶ On décide d'enlever toutes les données censurées : on étudie donc la loi de  $\{T, \Delta = 1\}$ .
- ▶ Classiquement, on suppose  $\tilde{T}$  **indépendante** de  $C$ .
- ▶ Pour la fonction de répartition des données non censurées, on a

$$\begin{aligned}\mathbb{P}[T \leq t, \Delta = 1] &= \mathbb{E}[I(\tilde{T} \leq t, \tilde{T} \leq C)] \\ &= \iint I(u \leq t, u \leq v) dF(u) dG(v) \\ &= \int I(u \leq t) \left( \int I(u \leq v) dG(v) \right) dF(u) \\ &= \int_0^t (1 - G(u)) dF(u)\end{aligned}$$

où  $F$  est la f.d.r de  $\tilde{T}$  et  $G$  la f.d.r de  $C$ .

## Approche naïve : justification mathématique

- ▶ On voit donc que :

$$\mathbb{P}[T \leq t, \Delta = 1] \neq \mathbb{P}[\tilde{T} \leq t] = \int_0^t dF(u),$$

sauf si  $G(u) = 0$ , ce que qui signifierait qu'il n'y a pas de censure !!

- ▶ Puisque  $0 \leq (1 - G(u)) \leq 1$ , on peut même dire que

$$\mathbb{P}[T \leq t, \Delta = 1] \leq \mathbb{P}[\tilde{T} \leq t].$$

En ne prenant que les données non censurées on a tendance à **sous évaluer** la f.d.r de la variable d'intérêt !

## Exemple sur des données simulées

```
TrueT<-runif(100,0,5)
Cens<-runif(100,0,4)
ObsT<-pmin(TrueT,Cens)
Delta<-TrueT<=Cens
```

##	TrueT	Cens	ObsT	Delta
## 1	1.61	3.38	1.61	1
## 2	4.47	2.12	2.12	0
## 3	4.43	1.39	1.39	0
## 4	0.85	3.40	0.85	1
## 5	1.75	1.39	1.39	0
## 6	0.51	1.36	0.51	1
## 7	1.29	3.99	1.29	1
## 8	2.41	0.98	0.98	0
## 9	3.58	0.34	0.34	0
## 10	2.44	2.79	2.44	1



## Exemple sur des données simulées

On compare les moyennes

- ▶ des données observées (les  $T_i$ )
- ▶ uniquement des données non censurées (les  $T_i$  tels que  $\Delta_i=1$ )
- ▶ des vraies durées (les  $\tilde{T}_i$ )

```
mean(ObsT)
```

```
## [1] 1.486174
```

```
mean(ObsT[Delta==1])
```

```
## [1] 1.318953
```

```
mean(TrueT)
```

```
## [1] 2.443657
```

Autres types de censure et troncature à gauche

## Définitions

Il existe d'autres types de données manquantes liées au suivi des patients dans le temps.

- ▶ la **censure à gauche** : on ne connaît pas toujours la date exacte d'entrée dans l'étude.
- ▶ la **censure par intervalle** : on ne connaît qu'un intervalle de temps par individu et on sait que l'évènement d'intérêt s'est produit dans cet intervalle.
- ▶ la **troncature à gauche** : l'entrée dans l'étude pour un patient est retardée ("delayed entry" en anglais) et les données sont recueillies conditionnellement au fait que l'évènement d'intérêt ne se soit pas produit avant que le suivi de l'individu ait commencé.

## Exemple de censure à gauche

Les babouins de la réserve d'Amboli, au Kenya, dorment dans les arbres et descendent de leurs arbres pour aller se nourrir. L'évènement d'intérêt est l'instant où ils descendent de l'arbre. Des biologistes viennent faire des visites régulières pour voir si les babouins sont descendus de leur arbre.

- ▶ L'évènement d'intérêt est observé si le babouin descend de l'arbre après l'arrivée des biologistes.
- ▶ Par contre, la donnée est censurée si le babouin est descendu avant l'arrivée des observateurs. On sait uniquement que l'horaire de descente est inférieur à l'heure d'arrivée des observateurs.
- ▶ On observe donc le maximum entre l'heure de descente des babouins et l'heure d'arrivée des observateurs.

## Exemple de censure par intervalles

On s'intéresse à des patients qui ont un plombage et qui vont régulièrement chez le dentiste. A chaque visite, le dentiste examine les dents du patient. L'évènement d'intérêt est le moment où le plombage tombe qui peut soit être dû à une fracture du plombage ou bien à une complication appelée complication endodontique.

- ▶ La fracture du plombage peut uniquement être observée par le dentiste et cette observation est donc **censurée par intervalle**. On sait seulement que la fracture a eu lieu entre les deux dernières visites chez le dentiste.
- ▶ Une complication endodontique s'accompagne d'une douleur intense et donc le patient connaît la date exacte où le plombage est tombé (**évènement d'intérêt observé**).
- ▶ Certains patients gardent leur plombage toute leur vie, et donc on n'observera jamais la date où le plombage est tombé. Ces observations sont **censurées à droite**.

## Exemple de troncature à gauche

Le “Steno memorial” est un hôpital Danois (à Copenhague) construit en 1933 et qui sert d'hôpital spécialisé sur le diabète pour tout le Danemark. On s'intéresse aux données de survie des patients diabétiques recueillis dans cet hôpital.

- ▶ L'évènement d'intérêt est la date de décès des patients.
- ▶ La variable d'intérêt est le temps écoulé depuis le diagnostic du diabète jusqu'au décès.
- ▶ Les patients ont été inclus dans l'étude de 1933 jusqu'à 1972.
- ▶ 2 709 patients ont été suivis depuis leur entrée à l'hôpital.
- ▶ L'étude s'est arrêtée le 31 décembre 1984.

## Exemple de troncature à gauche

- ▶ 707(26%) décès sont observés durant l'étude.
- ▶ Les 2 002(74%) patients restants ont été **censurés à droite** soit parce qu'ils ont quitté l'étude ou alors parce qu'ils étaient encore en vie au 31 décembre 1984.
- ▶ Les patients ne contactent pas l'hôpital tout de suite après avoir été diagnostiqués diabétiques. Un patient est donc observé **conditionnellement** à ce qu'il ne décède pas avant d'avoir contacté l'hôpital (en moyenne un patient met environ 6 ans avant de contacter l'hôpital).

Si on ne prend pas en compte la **troncature à gauche**, on va avoir tendance à sous évaluer la mortalité des diabétiques danois puisqu'on n'a pas les patients qui sont décédés avant d'avoir contacté l'hôpital !

# Les fonctions types utilisées en analyse de survie

- Introduction et propriétés



## Cinq fonctions essentielles ( $t \geq 0$ )

- ▶ la fonction de répartition :  $F(t) = \mathbb{P}[\tilde{T} \leq t]$ .
- ▶ la fonction de survie :  $S(t) = \mathbb{P}[\tilde{T} > t]$ .
- ▶ la densité :

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq \tilde{T} < t + \Delta t]}{\Delta t}$$

- ▶ le risque instantané ou “hazard rate” en anglais :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq \tilde{T} < t + \Delta t | \tilde{T} \geq t]}{\Delta t}$$

- ▶ le risque cumulé  $H(t)$  :

$$H(t) = \int_0^t h(s) ds.$$

## Cinq fonctions essentielles

- ▶  $F$  est une fonction **croissante** telle que  $F(0) = 0$  et  $\lim_{t \rightarrow \infty} F(t) = 1$ .
- ▶  $S$  est une fonction **décroissante** telle que  $S(0) = 1$  et  $\lim_{t \rightarrow \infty} S(t) = 0$ . On a la relation :  $S(t) = 1 - F(t)$ .
- ▶ la densité mesure la probabilité que l'évènement d'intérêt se produise sur un intervalle de temps infinitésimal. On a les relations :  $F(t) = \int_0^t f(x)dx$  et  $F'(t) = f(t)$ .
- ▶ le risque instantané mesure la probabilité que l'évènement d'intérêt se produise sur un intervalle de temps infinitésimal ( $[t, t + \Delta t]$ ), conditionnellement au fait d'**être à risque** que l'évènement se produise ( $\tilde{T} \geq t$ ).

Si le temps est en jours :  $h(t)$  est la probabilité de décéder le jour  $t + 1$  sachant que l'on est encore en vie au jour  $t$ .

## Relations entre risque instantané et survie (1)

D'après la formule de Bayes :

$$\begin{aligned}h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq \tilde{T} < t + \Delta t | \tilde{T} \geq t]}{\Delta t} \\&= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq \tilde{T} < t + \Delta t, \tilde{T} \geq t]}{\Delta t} \times \frac{1}{P[\tilde{T} \geq t]} \\&= \frac{f(t)}{S(t)}.\end{aligned}$$

Puisque  $S'(t) = (1 - F(t))' = -f(t)$ , on a également :

$$h(t) = -\frac{S'(t)}{S(t)}.$$

## Relations entre risque instantané et survie (2)

Par ailleurs  $S'(t)/S(t) = [-\log(S(t))]'$  et donc

$$\begin{aligned}h(t) &= [-\log(S(t))]' \\ \int_0^t h(u) du &= [-\log(S(t))] \\ \exp\left(-\int_0^t h(u) du\right) &= S(t).\end{aligned}$$

La f.d.r, la fonction de survie, la densité, le risque instantané et le risque cumulé caractérisent tous la loi de  $\tilde{T}$ , de telle sorte que si l'on connaît l'une de ces cinq fonctions on peut retrouver les quatre autres !

## Exemples de risques instantanés

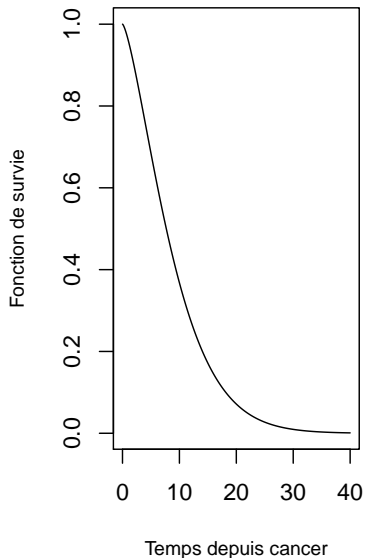
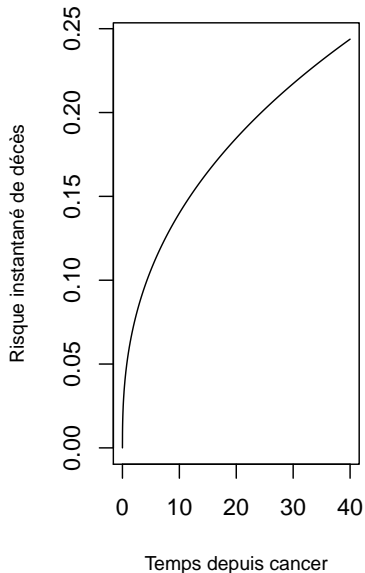
- ▶ le risque instantané **constant** :  $h(t) = a$ . On a alors  $S(t) = \exp(-at)$ ,  $\tilde{T} \sim \mathcal{E}(a)$ . C'est une loi dite sans mémoire ! Loi typique pour étudier la durée de vie de composants électroniques, d'une ampoule etc.
- ▶ le risque instantané **croissant** : il y a "vieillessement" ! C'est le risque le plus classique. Par exemple, pour modéliser la durée de vie de patients à partir du moment qu'ils ont développé une maladie pulmonaire chronique, qu'ils ont eu un certain cancer etc. Egalement, la durée de vie d'une voiture à partir du premier accident ...
- ▶ le risque instantané **décroissant** : assez rare en pratique, révèle généralement de l'hétérogénéité dans la survie des patients. Un exemple classique est la mortalité infantile jusqu'à un an dans un pays "pauvre".

## Exemples de risque instantané croissant/décroissant

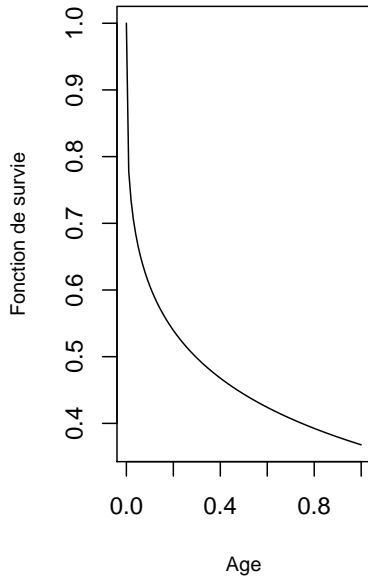
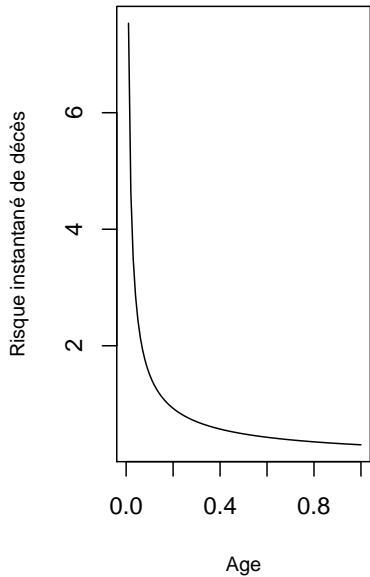
On peut modéliser un risque instantané monotone croissant/décroissant par une loi de Weibull de paramètre de forme  $a$  ( $a > 1$  risque croissant,  $a < 1$  risque décroissant) et de paramètre d'échelle  $b$ .

$$h(t) = \frac{a}{b} \left(\frac{t}{b}\right)^{a-1}$$
$$S(t) = \exp\left(-\left(\frac{t}{b}\right)^a\right).$$

## Exemples de risque instantané croissant ( $a = 1.4, b = 10$ )



## Exemples de risque instantané décroissant ( $a = 0.3, b = 1$ )





## Exemples de risques instantanés

- ▶ Il existe aussi des risques instantanés en forme de  $\cup$  (“bathtub” en anglais) : risque diminue en début, se stabilise ensuite puis risque élevé en fin de vie.
- ▶ Enfin, il existe des risques instantanés en forme de  $\cap$  : risque élevé en début, se stabilise ensuite puis risque faible en fin de vie.

## Quelques propriétés sur la fonction de survie

- ▶ Dans le chapitre suivant on verra comment estimer le risque instantané, sans introduire de **biais**, à partir des observations :

$$\begin{cases} T_i = \min(\tilde{T}_i, C_i) \\ \Delta_i = I(\tilde{T}_i \leq C_i). \end{cases}$$

- ▶ Une fois le risque instantané on pourra estimer la fonction de survie.
- ▶ A partir de cet estimateur on essaiera de retrouver des quantités d'intérêt comme la médiane de  $\tilde{T}$ , les quartiles, la moyenne, la variance etc.

## Quelques propriétés sur la fonction de survie

- Pour l'espérance et la variance, on montre par le théorème de Fubini que :

$$\mathbb{E}[\tilde{T}] = \int_0^{\infty} S(t) dt$$

$$\mathbb{V}[\tilde{T}] = 2 \int_0^{\infty} S(t) dt - \left( \int_0^{\infty} S(t) dt \right)^2.$$

- Le quantile d'ordre  $p$  de  $\tilde{T}$ ,  $Q(p)$ , est égal à :

$$\begin{aligned} Q(p) &= \inf\{t : F(t) \geq p\} = \inf\{t : 1 - S(t) \geq p\} \\ &= \inf\{t : S(t) \leq 1 - p\}. \end{aligned}$$

## Exemples d'études de survie

# Etude sur l'association entre accouchement par césarienne et diabète de type 1

ORIGINAL ARTICLE

## Prelabor Cesarean Section and Risk of Childhood Type 1 Diabetes A Nationwide Register-based Cohort Study

Tine Dalsgaard Clausen,<sup>a</sup> Thomas Bergholt,<sup>a</sup> Frank Eriksson,<sup>b</sup> Steen Rasmussen,<sup>c</sup>  
Niels Keiding,<sup>b</sup> and Ellen C. Løkkegaard<sup>a</sup>

**Background:** Unfavorable conditions associated with cesarean section may influence the risk of type 1 diabetes in offspring, but results from studies are conflicting. We aimed to evaluate the association between prelabor cesarean section and risk of childhood type 1 diabetes.

**Methods:** A Danish nationwide cohort study followed all singletons born during 1982–2010. Four national registers provided information on mode of delivery, outcome, and confounders. The risk of childhood type 1 diabetes with onset before the age of 15 years was assessed by Cox regression. A total of 1,760,336 singletons contributed 20,356,684 person-years, during which 4,400 were diagnosed with childhood type 1 diabetes.

**Results:** The hazard ratio (HR) for childhood type 1 diabetes was increased in children delivered by prelabor cesarean section compared with vaginal delivery when adjusted for year of birth, parity, sex, parental age, and education and paternal type 1 diabetes status at childbirth (HR = 1.2; 95% confidence interval [CI] = 1.0, 1.3), but not after additional adjustment for maternal type 1 diabetes status at childbirth (HR = 1.1; 95% CI = 0.95, 1.2). Delivery by intrapartum cesarean section was not associated with childhood type 1 diabetes.

Paternal type 1 diabetes was a stronger risk factor for childhood type 1 (HR = 1.2; 95% CI = 1.0, 1.4) than maternal type 1 diabetes (HR = 0.65; 95% CI = 0.52, 0.80).

**Conclusions:** Delivery by prelabor cesarean section was not associated with an increased risk of childhood type 1 diabetes in the offspring.

(*Epidemiology* 2016;27: 547–555)

During the past decades, incidence rates of type 1 diabetes have increased rapidly, not only in high-income countries or in populations with a high genetic predisposition, but also globally.<sup>1–6</sup> Among 0- to 14-year-old children, incidence rates between 0.1 and 60 per 100,000 person-years have been observed, with highest rates in Finland and lowest rates in East Asian populations.<sup>5,7</sup> The steep rise cannot be explained by genes alone, and effort has been made to isolate modifiable pathogenic, epigenetic, and environmental factors.

The autoimmune response and maturation of the immune system may be modulated by perinatal events related to intrapartum stress response, anesthetic agents, and bacterial colonization of the neonate intestine. It has been observed that children with type 1 diabetes present a gut microbiota, which is less diverse and contains more *Clostridium* species than children without type 1 diabetes; it has also been suggested that the bacterial composition may correlate with blood glucose levels.<sup>8</sup> It has furthermore been found that children delivered by cesarean section present a gut microbiota, which resembles the one found in children with type 1 diabetes,<sup>9,10</sup> but results from studies looking at a potential association between delivery by cesarean section and type 1 diabetes in the offspring are divergent.<sup>11–17</sup> The combination of steadily increasing rates of cesarean section reported in most countries, as well as the expanding research that suggest a serious influence of gut microbiota on future health and disease, emphasize the need for large-scale studies to evaluate the clinical consequences of delivery by cesarean section in the general population.<sup>12,18</sup>

Submitted 3 June 2015; accepted 24 March 2016.

From the Department of Gynecology and Obstetrics, Nordsjællands Hospital, University of Copenhagen, Hillerød, Denmark; <sup>a</sup>Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark; and <sup>b</sup>Department of Microbiology, Hvidovre Hospital, Copenhagen, Denmark.

Supported by grants from Nordsjællands Hospital, Hillerød, University of Copenhagen, Denmark (No. 473). Nordsjællands Hospital had no role regarding design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

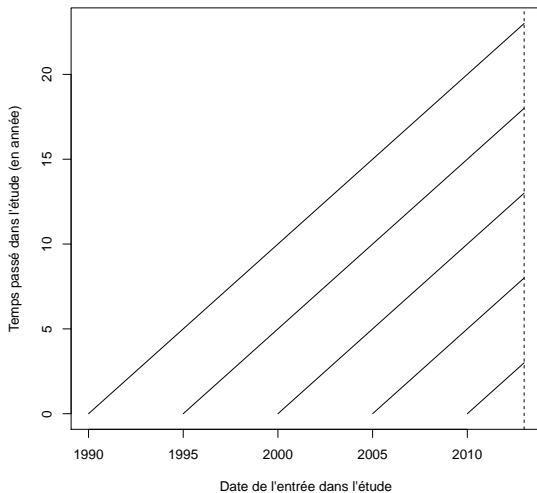
Preliminary results from this study were presented in abstract form at the Diabetes in Pregnancy Study Group of the European Association for the Study of Diabetes, Budapest, Hungary, 4 October 2014, and as a Guided Audio Poster Tour and in abstract form at the American Diabetes Association's 75th Scientific Sessions, Boston, MA, June 2015.

The authors report no conflicts of interest.

Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article ([www.epidem.com](http://www.epidem.com)).

Correspondence: Tine Dalsgaard Clausen, Department of Gynecology and Obstetrics, Nordsjællands Hospital, Dyrhøveløvej 29, 3400 Hillerød, Denmark. E-mail: [tine.dalsgaard.clausen@slhdtinet.dk](mailto:tine.dalsgaard.clausen@slhdtinet.dk).

# Le diagramme de Lexis



## Etude sur la durée de vie des gauchers

Les droitiers vivraient 9 ans de plus en moyenne que les gauchers d'après une étude chez Nature (1980) et une autre étude chez Science (1990) ...

<http://www.bbc.com/news/magazine-23988352>