Epidémiologie

Plan

Philippe SAINT PIERRE

Université Paul Sabatier – Toulouse III

Institut de Mathématiques de Toulouse

philippe.saint-pierre@math.univ-toulouse.fr

Epidémiologie

4. Mesures de risques

Philippe SAINT PIERRE

Université Paul Sabatier – Toulouse III

Institut de Mathématiques de Toulouse

philippe.saint-pierre@math.univ-toulouse.fr

4. Mesures de risques

I. Prévalence

II. Taux d'incidence

- Personnes-temps
- Enquêtes de cohorte
- Enquêtes transversale répétées
- Données de surveillance épidémiologique

III. Risque de la maladie

- TI constant
- TI variables

I. Prévalence

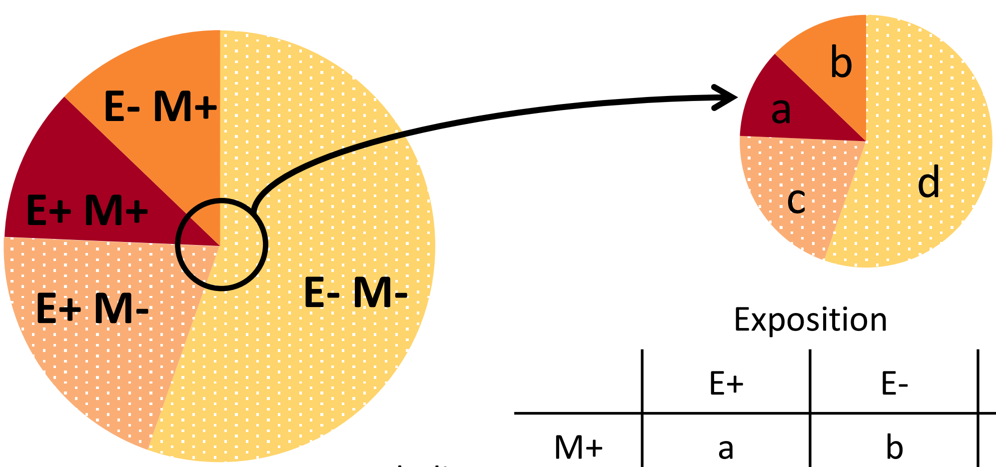
- Proportion de malades M+ dans une population de taille N, à un instant donné

$$\boxed{P = \frac{M_1}{N} = \frac{M_1}{M_0 + M_1}}$$

- Probabilité : $P \in [0,1]$
- Estimation dans les enquêtes transversales (échantillon représentatif)

Population

Echantillon



Exposition

Maladie	E+	E-	
M+	a	b	m_1
M-	c	d	m_0
	n_1	n_0	n

Prévalence de la maladie

$$\hat{P}(M+) = \frac{m_1}{n} = \frac{a+b}{n}$$

Prévalence de l'exposition

$$\hat{P}(E+) = \frac{n_1}{n} = \frac{a+c}{n}$$

Prévalence

- Prévalence de la maladie $$\boxed{P(M+) = \frac{M_1}{N} = \frac{M_1}{M_0+M_1}}$$

- Estimation de la prévalence dans un échantillon représentatif $$\boxed{\hat{P} = \frac{m_1}{n}}$$

- Intervalle de confiance de niveau α

$$\boxed{IC_\alpha(P) = [P_{inf}\,; P_{sup}] = \left[\hat{P} - z_{\frac{\alpha}{2}}\sqrt{Var(\hat{P})}\,;\ \hat{P} + z_{\frac{\alpha}{2}}\sqrt{Var(\hat{P})}\right] \quad \text{avec} \quad Var(\hat{P}) = \frac{\hat{P}(1-\hat{P})}{n}}$$

Conditions d'applications

- approximation de la loi Normale par la loi Binomiale : $B(n,p) \xrightarrow[n\to\infty]{} N(np, np(1-p))$
- en pratique, $nP_{inf} \geq 5, nP_{sup} \geq 5, nQ_{inf} \geq 5, nQ_{sup} \geq 5$ $\qquad (Q = 1 - P)$

Prévalence : exemple

- Echantillon de 100 travailleurs d'une usine

Cancer	M+	20
	M-	80
		100

- Estimation de la prévalence $\hat{P} = \frac{m_1}{n} = \frac{20}{100} = 0.2$

- Intervalle de confiance à 95%

$$IC_{0.05}(P) = \left[0.2 \pm 1.96\sqrt{\frac{0.2 \times (1-0.8)}{100}}\right] = [0.12\,;0.28]$$

- Conditions d'applications :
  - $100 \times 0.12 = 12 \geq 5;\ 100 \times 0.28 = 28$
  - $100 \times 0{,}88 = 88 \quad \geq 5;\ \ 100 \times 0{,}72 = 72$

Prévalence

- Pas toujours un bon indicateur :
  - dépend de la durée de la maladie

    durée augmente → prévalence augmente

  - dépend de la vitesse d'apparition des nouveaux cas de la maladie

    vitesse augmente → prévalence augmente

  - Ex : prévalence d'une maladie chronique et d'une infection aigue
  - Ex : augmentation de la prévalence lié à des améliorations de santé

- Photographie à un moment donné : introduction d'un nouvel indicateur

→ Taux d'incidence

II. Taux d'incidence

- Taux d'incidence représente la vitesse d'apparition de nouveaux (vitesse moyenne)

$$TI = \frac{I}{PT} = \frac{Nbre\ de\ nouveaux\ cas\ de\ la\ maladie}{Nbre\ de\ "personnes - temps"}$$

- PT est la somme des temps d'exposition de chaque sujet pendant la période de suivi
  - Ex: 20 sujets suivis pendant 2 ans correspondent à 40 personnes-année

- Unité du taux d'incidence : nombre de nouveaux cas par $"personnes - temps"$

- Intervalle de confiance de niveau α

$$IC_{\alpha}(TI) = [TI_{inf} ; TI_{sup}] = \left[\widehat{TI} - z_{\frac{\alpha}{2}}\sqrt{Var(\widehat{TI})} ; \widehat{TI} + z_{\frac{\alpha}{2}}\sqrt{Var(\widehat{TI})}\right]$$

avec $Var(\widehat{TI}) = \frac{I}{PT^2}$

Taux d'incidence : personnes-temps

- Sujet à risque : sujet non malade susceptible de devenir malade
- "$personnes - temps$" d'un sujet à risque
  - Temps d'exposition = durée pendant laquelle un sujet peut être enregistré comme cas s'il développe la maladie étudiée
  - Le nombre de "$personnes - temps$" est spécifique à chaque individu et dépend de
    - la date de début du suivi : t_{0i}
    - la date de fin du suivi : t_{fi}

$$\boxed{PT_i = \Delta t_i = t_{0i} - t_{fi}}$$

- La date de fin de suivi peut être due à
  - La survenue de la maladie M
  - La censure à droite
    - Exclu-vivant : fin de l'étude et du suivi
    - Perdu de vue (déménagement, décès, arrêt volontaire du suivi, …)

Taux d’incidence : estimation

- Estimation du taux d’incidence

a. Les enquêtes de cohorte

→ population fermée et suivi individuel des sujets

a. Les enquêtes transversales répétées

→ population fermée mais dates d’évènement inconnues

a. Les données de surveillance épidémiologique

→ population ouverte et données groupées

Taux d'incidence : enquête de cohorte

a) Taux d'incidence à partir d'une enquête de cohorte

- On connait les temps d'exposition individuel

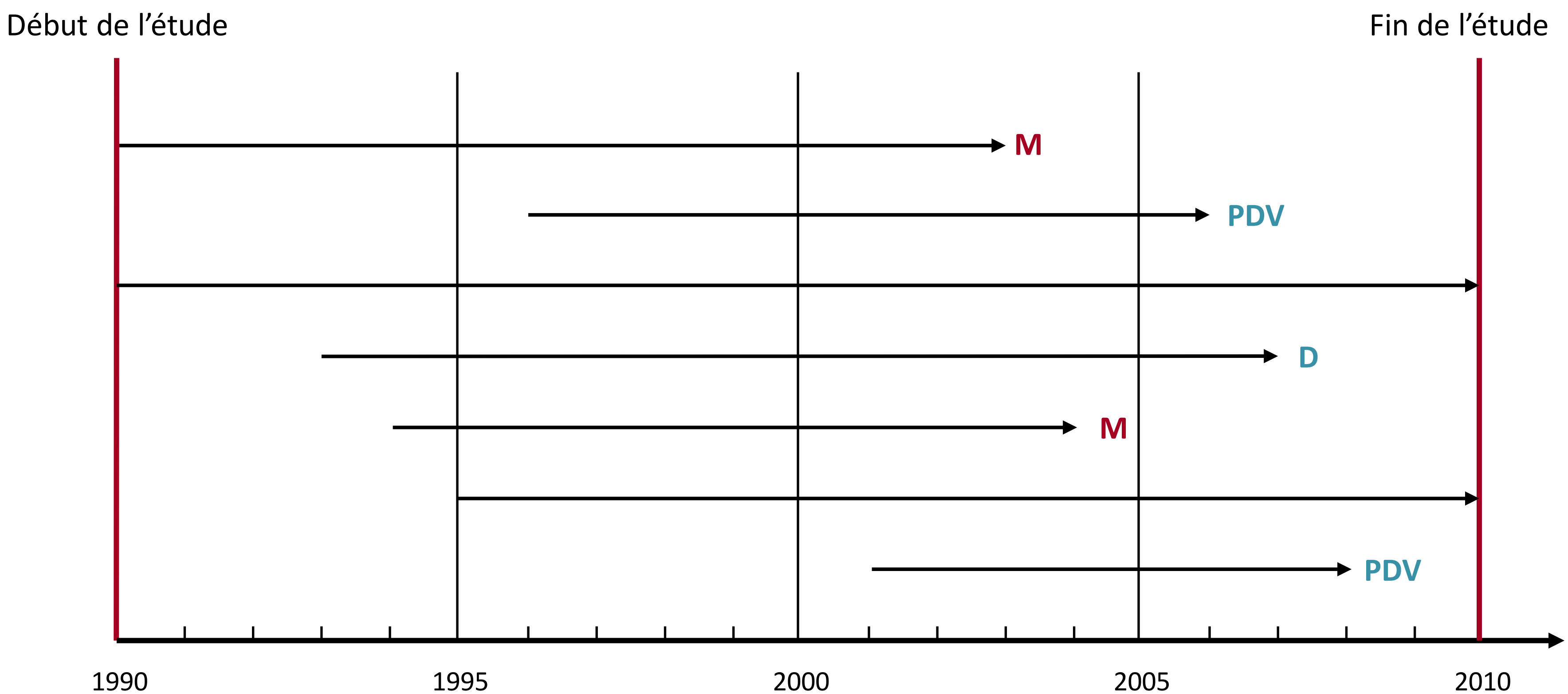
$$\boxed{\begin{array}{c} \Delta t_i = t_{0i} - t_{fi} \\ PT = \sum_{i=1}^{n} \Delta t_i \end{array}}$$

- On connait le nombre de nouveau cas I sur la période de suivi

- Le taux d'incidence peut être estimé par

$$\boxed{\widehat{TI} = \frac{I}{\sum_{i=1}^{n} \Delta t_i}}$$

Taux d'incidence : enquête de cohorte



- $PT = 13 + 10 + 20 + 14 + 10 + 15 + 7 = 89$ *personnes − année*
- 2 malades pendant la période de l'étude (1990-2010)
- Taux d'incidence entre 1990 et 2010 : $TI = \frac{2}{89} = 0.0224\ cas\ pour\ 1\ personne - année$

Taux d'incidence : enquête de cohorte

- $PT = 89\ personnes - année$

  $= 89\ sujets\ suivi\ pendant\ 1\ année$

  $= 1068\ sujet\ suivi\ pendant\ 1\ mois$

  $= 1\ sujet\ suivi\ pendant\ 89\ années$

  $= 0.89\ sujet\ suivi\ pendant\ 1\ siècle$

- $TI = \frac{2}{89} = 0.0224\ cas\ pour\ 1\ personne - année = 0{,}0224\ cas/PA$

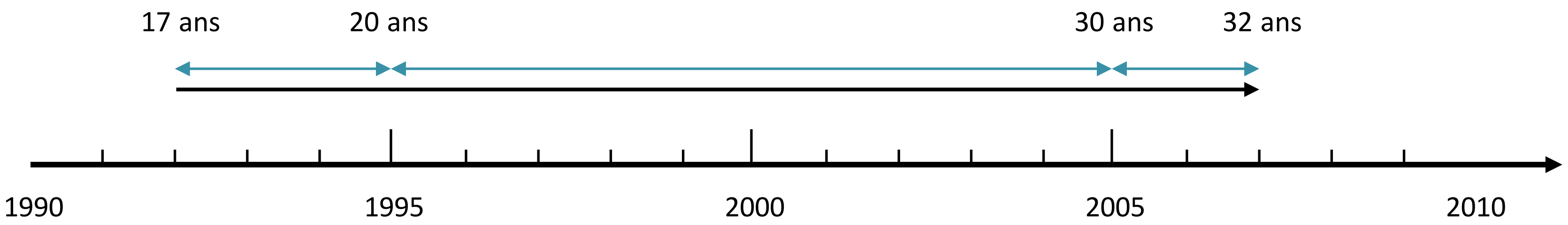
  $= 22{,}4\ cas\ pour\ 1000\ personnes - année$

- $TI = \frac{2}{1068} = 0.0019\ cas\ pour\ 1\ personne - mois = 0{,}0019\ cas/PM$

  $= 1.9\ cas\ pour\ 1000\ personnes - mois$

Taux d’incidence : enquête de cohorte

- Calcul des personnes-temps par intervalle de temps



Nombre de personnes-année par tranche d’âge

- $Age \in [15-20[\;\rightarrow 3 \; PA$
- $Age \in [20-30[\;\rightarrow 10 \; PA$
- $Age \in [30-40[\;\rightarrow 2 \; PA$

- Taux d’incidence en fonction de l’âge (Maladie coronarienne et contraception orale)

Age	I	PA	TI $\times 10^3$
$[15-20[$	40	9228	4.3
$[20-30[$	88	8376	10.5
$[30-40[$	130	7092	18.3
	258	24696	10.4

Augmentation du taux d’incidence avec l’âge

Taux d'incidence : enquêtes transversales répétées

b) Taux d'incidence à partir d'une enquête transversale recontactée après une période Δt

- On dispose de l'informations sur
  - Le nombre de sujets N inclus dans l'étude et recontactés après une période Δt
  - Le nombre de perdus de vue PDV sur la période Δt
  - Le nombre de décès D sur la période Δt
  - Le nombre sujets I qui ont développés la maladie sur la période Δt

- Le nombre de personnes-temps peut être estimé par

$$\boxed{\widehat{PT} = N \times \Delta t - \left[(I + PDV + D) \times \frac{\Delta t}{2}\right] = \left[N - \frac{(I + PDV + D)}{2}\right] \times \Delta t}$$

  - en supposant que la survenue des décès, des cas et des PDV est uniforme, c'est-à-dire qu'elle se produit en moyenne au milieu de l'intervalle

- Le taux d'incidence : $\boxed{\widehat{TI} = \frac{I}{\widehat{PT}}}$

Taux d'incidence : données de surveillance

c) Taux d'incidence à partir des <u>données de surveillance épidémiologique</u>

- Population ouverte (Ex: une ville, une région, …), on dispose de l'informations sur
  - Une estimation du nombre moyen de sujets N de la population sur la période Δt

    ↳ obtenu à partir d'un ou plusieurs recensements ou estimations intercensitaires
  - Le nombre sujets I qui ont développés la maladie sur la période Δt

    ↳ obtenu à partir d'un système d'enregistrement des cas (registre, …)
- On suppose un renouvellement constant de la population (autant d'entrées que de sorties : naissances, décès, immigrations, émigrations)

  ➡ Le nombre de personnes-temps peut être estimé par $\boxed{\widehat{PT} = N \times \Delta t}$
- En supposant que la survenue de la maladie est uniforme, le taux d'incidence peut être estimé

$$\boxed{\widehat{TI} = \frac{I}{\widehat{PT}}}$$

III. Risque de la maladie

- Le risque de la maladie représente la probabilité de devenir malade au cours d'une période de temps Δt

- Soit X le temps écoulé avant l'apparition de la maladie

$$R(\Delta t) = P(X < t) = 1 - P(X \geq t) = 1 - \exp\left(-\int_0^{\Delta t} TI \, dt\right)$$

$$\boxed{R(\Delta t) = 1 - \exp(-TI \times \Delta t)}$$

- Le risque sur une période Δt peut être estimé par

$$\boxed{\hat{R}(\Delta t) = 1 - \exp\left(-\widehat{TI} \times \Delta t\right)}$$

- Hypothèse : TI est constant sur la période Δt

Risque de la maladie

- Intervalle de confiance de niveau α de R

$$\boxed{IC_\alpha(R) = \left[R_{inf}\,; R_{sup}\right] = \left[1 - \exp(-TI_{inf} \times \Delta t)\,; 1 - \exp(-TI_{sup} \times \Delta t)\right]}$$

$$IC_\alpha(TI) = \left[TI_{inf}\,; TI_{sup}\right] = \left[\widehat{TI} - z_{\frac{\alpha}{2}}\sqrt{Var\left(\widehat{TI}\right)}\,;\ \hat{P} + z_{\frac{\alpha}{2}}\sqrt{Var\left(\widehat{TI}\right)}\right] \quad \text{avec} \quad Var\left(\widehat{TI}\right) = \frac{I}{PT^2}$$

- Si $\widehat{TI} \times \Delta t$ est petit alors $\hat{R}(\Delta t) \approx \widehat{TI} \times \Delta t$

- Remarque : concordance des unités
  - Si Δt en années utiliser un TI annuel (cas/PA)
  - Si Δt en mois utiliser un TI mensuel (cas/PM)

Risque de la maladie : exemple

- 1200 sujets inclus dans une étude
- Au bout de 5 ans, on observe 65 cas de maladie, 50 perdus de vue et 10 décès

- Taux d'incidence TI

$$\widehat{TI} = \frac{I}{PT} = \frac{65}{\left[1200 - \frac{50+10+65}{2}\right] \times 5} = 0.0114 = 114\ cas\ pour\ 10\ 000\ PA$$

$$IC_{95\%}(TI) = \left[TI_{inf}\ ;TI_{sup}\right] = \left[\widehat{TI} - 1.96\sqrt{\frac{I}{PT^2}}\ ;\ \widehat{TI} + 1.96\sqrt{\frac{I}{PT^2}}\right] = [0.086\ ; 0.142]$$

- Risque de développer la maladie dans les 5 ans

$$\hat{R}(\Delta t) = 1 - \exp\left(-\widehat{TI} \times \Delta t\right) = 1 - \exp(-0.0114 \times 5) = 5.54 \times 10^{-2} \longrightarrow 5.54\%$$

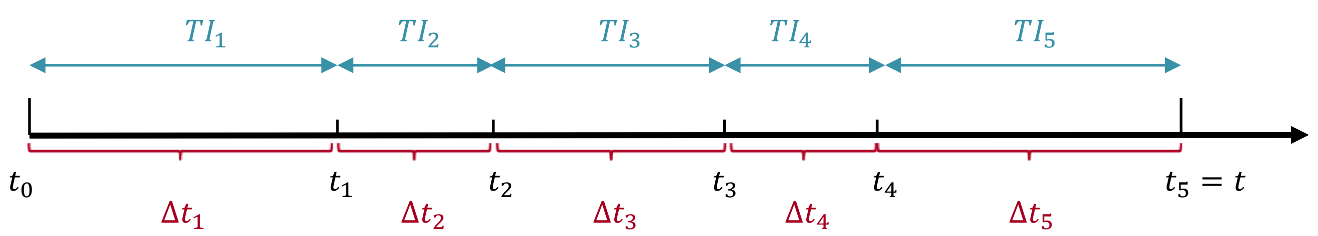
$$IC_{95\%}(R) = \left[1 - \exp(-TI_{inf})\ ; 1 - \exp(-TI_{sup})\right] = [4.2 \times 10^{-2}\ ; 6.9 \times 10^{-2}]$$

- Risque de développer la maladie dans les 3 ans

$$\hat{R}(\Delta t) = 1 - \exp\left(-\widehat{TI} \times \Delta t\right) = 1 - \exp(-0.0114 \times 3) = 3.4 \times 10^{-2} \longrightarrow 3.4\%$$

Risque de la maladie : TI variables

Taux d'incidence variables sur la période Δt



$\Delta t = t - t_0$

$\Delta t_i = t_i - t_{i-1}$

$\Delta t = \sum_k \Delta t_k$

- Soit TI_i le taux d'incidence sur $[t_{i-1}, t_i[$, TI_i supposé constant sur $[t_{i-1}, t_i[$

- Le risque sur une période Δt est

$$\boxed{R(\Delta t) = 1 - \exp\left(-\sum_k TI_k \times \Delta t_k\right)}$$

- Le risque sur une période Δt peut être estimé par

$$\boxed{\hat{R}(\Delta t) = 1 - \exp\left(-\sum_k \widehat{TI}_k \times \Delta t_k\right)}$$

Risque de la maladie : TI variables

- Intervalle de confiance de niveau α de A $= \sum_k TI_k \times \Delta t_k$

$$\boxed{IC_\alpha(A) = \left[A_{inf}\,;A_{sup}\right] = \left[\hat{A} - z_{\frac{\alpha}{2}}\sqrt{Var(\hat{A})}\,;\hat{A} + z_{\frac{\alpha}{2}}\sqrt{Var(\hat{A})}\right]}$$

$$\widehat{A} = \sum_k \widehat{TI}_k \times \Delta t_k \quad \text{et} \quad Var\left(\hat{A}\right) = \sum_k Var\left(\widehat{TI}_k\right) \times {\Delta t_k}^2 \boxed{= \sum_k \frac{I_k}{{PT_k}^2} \times {\Delta t_k}^2}$$

- Intervalle de confiance de niveau α de R

$$\boxed{IC_\alpha(R) = \left[R_{inf}\,;R_{sup}\right] = \left[1 - \exp(-A_{inf})\,;1 - \exp(-A_{sup})\right]}$$

- Si $\widehat{TI}_k \times \Delta t_k$ est petit alors $\hat{R}(\Delta t) \approx \sum_k \widehat{TI}_k \times \Delta t_k$

- Remarque : concordance des unités

Risque de la maladie : TI variables

- Décès par maladie coronarienne : 343 hommes d'une entreprise de sulfure de carbone suivis pendant 10 ans

Age	Décès	PA	TI pour 1000 PA
$[25-40[$	0	480	0
$[40-45[$	1	587	1.7
$[45-50[$	3	680	4.4
$[50-55[$	5	541	9.2
$[55-60[$	8	479	16.7
$[60-65[$	8	356	22.5
$[65-70[$	4	157	25.5
	29	3280	8.8

- Risque de décéder pendant une période donnée ?

- Hypothèses : - TI constant dans chaque tranche d'âge

  - Pas d'effet génération (risque de décéder entre 50 et 55 ans est une bonne estimation pour un individu de 30 ans)

Risque de la maladie : TI variables

- Risque de décéder entre 25 et 55 ans

$$R(25-55) = 1 - \exp(-\sum_k TI_k \times \Delta t_k)$$
$$= 1 - \exp(-[0 \times 15 + 5 \times (0.0017 + 0.0044 + 0.0092)])$$
$$= 1 - \exp(-0.0765) = 0,0736$$

Le risque de décéder entre 25 et 75 ans est de 7.36%

- Intervalle de confiance

  - $\widehat{A} = \sum_k \widehat{TI}_k \times \Delta t_k = 0.0765$
  - $Var(\hat{A}) = \sum_k \frac{I_k}{{PT_k}^2} \times {\Delta t_k}^2 = \frac{0}{480^2} \times 15^2 + \frac{1}{587^2} \times 5^2 + \frac{3}{680^2} \times 5^2 + \frac{5}{541^2} \times 5^2 = 0.0007$

$$IC_{95\%}(A) = \left[A_{inf} \,;\, A_{sup}\right] = \left[0.0765 \pm 1.96\sqrt{0.0007}\right] = [0.026 \,;\, 0.127]$$

$$IC_{95\%}(R) = [1 - \exp(-0.026) \,;\, 1 - \exp(-0.127)]$$

$$IC_{95\%}(R) = [0.025 \,;\, 0.119]$$

Taux d'incidence et prévalence

- Relation entre le taux d'incidence et la prévalence

$$\boxed{P = \frac{TI \times d}{1 + TI \times d}}$$

  - d est la durée moyenne de la maladie
  - TI le taux d'incidence
  - P la prévalence

- Hypothèse : population stable

- Si TI est petit alors $P \approx TI \times d$

Epidémiologie

5. Mesures d'association

Philippe SAINT PIERRE

Université Paul Sabatier – Toulouse III

Institut de Mathématiques de Toulouse

philippe.saint-pierre@math.univ-toulouse.fr

5. Mesures d’association

I. Mesures d’association

- Estimation du risque relatif et de l’odds ratio
- Intervalles de confiance et test d’une association
- Mesures d’association et types d’enquêtes

II. Notion de causalité

- Critères de Bradford Hill

III. Mesures d’impact potentiel

- Intervalles de confiance

		Exposition		
		E+	E-	
Maladie	M+	A	B	M_1
	M-	C	D	M_0
		N_1	N_0	

- Exposition est t'elle associée à la maladie ?

  ⟶ Mesures d'association : Risque relatif, Odds ratio

- Est-ce une relation causale ?

  ⟶ Critères de causalité, faisceau d'arguments

- Si relation causale, quelle est la proportion de cas attribuable à l'exposition ?

  ⟶ Mesures d'impact potentiel

I. Mesures d'association

	Exposition		
	E+	E-	
M+	A	B	M_1
M-	C	D	M_0
	N_1	N_0	

Maladie

$$P_1 = P(M + | E +) = \frac{A}{N_1}$$

$$P_0 = P(M + | E -) = \frac{B}{N_0}$$

- Modèle additif $\longrightarrow$ Différence de risque : $\Delta = P_1 - P_0$
- Modèle multiplicatif $\longrightarrow$ Risque relatif : $RR = \frac{P_1}{P_0}$

$\longrightarrow$ Odds ratio : $OR = \frac{P_1/(1-P_1)}{P_0/(1-P_0)}$

- Absence d'association : $\Delta = 0$ $RR = 1$ $OR = 1$

Estimation des mesures d'association

		Exposition		
		E+	E-	
Maladie	M+	a	b	m_1
	M-	c	d	m_0
		n_1	n_0	n

$$\hat{P}_1 = \hat{P}(M + | E +) = \frac{a}{n_1}$$

$$\hat{P}_0 = \hat{P}(M + | E -) = \frac{b}{n_0}$$

- Différence de risque : $\hat{\Delta} = \hat{P}_1 - \hat{P}_0 = \frac{a}{n_1} - \frac{b}{n_0}$
- Risque relatif : $\widehat{RR} = \frac{\hat{P}_1}{\hat{P}_0} = \frac{a/n_1}{b/n_0}$
- Odds ratio : $\widehat{OR} = \frac{\hat{P}_1/(1-\hat{P}_1)}{\hat{P}_0/(1-\hat{P}_0)} = \frac{ad}{bc}$

Exemple 1

		Alcool (consommation quotidienne)		
		≥ 40 cl	<40 cl	
Hypotrophie fœtale	M+	20	189	209
	M-	392	7372	7764
		412	7561	

$$\widehat{RR} = \frac{20/412}{189/7561} = 1.94$$

$$\widehat{OR} = \frac{20 \times 7372}{189 \times 392} = 1.99$$

Exemple 2

		Accouchement prématuré				
		M-	M+	P(M+|E)	$\widehat{RR}$	$\widehat{OR}$
Durée de scolarisation des parents	> 12 ans	243	7	2.8%	1	1
	9 - 12 ans	412	16	3.7%	1.34	1.35
	≤ 9 ans	181	9	4.7%	1.69	1.73

	9 - 12 ans	> 12 ans
M+	16	7
M-	412	243
	428	250

$$\widehat{RR} = \frac{16/428}{7/250} = \frac{3.7}{2.8} = 1.34 \quad ; \quad \widehat{OR} = \frac{16 \times 243}{412 \times 7} = 1.35$$

	≤ 9 ans	> 12 ans
M+	9	7
M-	181	243
	190	250

$$\widehat{RR} = \frac{9/190}{7/250} = \frac{4.7}{2.8} = 1{,}69 \quad ; \quad \widehat{OR} = \frac{9 \times 243}{181 \times 7} = 1.73$$

Intervalles de confiance

	E+	E-	
M+	a	b	m_1
M-	c	d	m_0
	n_1	n_0	n

$$\widehat{RR} = \frac{a/n_1}{b/n_0}$$

$$\widehat{OR} = \frac{ad}{bc}$$

- $\ln(\widehat{RR})$ et $\ln(\widehat{OR})$ convergent vers des lois normales quand $n \to \infty$

$$IC_\alpha(\ln(RR)) \coloneqq [B_{inf}; B_{sup}] = \ln(\widehat{RR}) \pm z_{\frac{\alpha}{2}}\sqrt{Var\left(\ln(\widehat{RR})\right)} \quad \text{et} \quad Var\left(\ln(\widehat{RR})\right) = \frac{c}{a \times n_1} + \frac{d}{b \times n_0}$$

$$IC_\alpha(\ln(OR)) \coloneqq [L_{inf}; L_{sup}] = \ln(\widehat{OR}) \pm z_{\frac{\alpha}{2}}\sqrt{Var\left(\ln(\widehat{OR})\right)} \quad \text{et} \quad Var\left(\ln(\widehat{OR})\right) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

- Intervalle de confiance de niveau α pour RR et OR

$$IC_\alpha(RR) \coloneqq [e^{B_{inf}}; e^{B_{sup}}]$$

$$IC_\alpha(OR) \coloneqq [e^{L_{inf}}; e^{L_{sup}}]$$

Exemple 1 (suite)

		Alcool (consommation quotidienne) ≥ 40 cl	<40 cl	
Hypotrophie fœtale	M+	20	189	209
	M-	392	7372	7764
		412	7561	

$$\widehat{RR} = \frac{20/412}{189/7561} = 1.94\,; \quad \ln(\widehat{RR}) = 0.67$$

$$Var(\ln(\widehat{RR})) = \frac{392}{20 \times 412} + \frac{7372}{189 \times 7561}$$
$$= 0.053$$

$$IC_{95\%}(\ln(RR)) \coloneqq 0.67 \pm 1.96\sqrt{0.053}$$
$$[\ 0.219\,;\, 1.121\ ]$$

$$IC_{95\%}(RR) \coloneqq \quad [\ e^{0.219}\,;\, e^{1.121}\ ]$$
$$[\ 1.25\,;\ 3.07\ ]$$

$$\widehat{OR} = \frac{20 \times 7372}{189 \times 392} = 1.99\,; \quad \ln(\widehat{OR}) = 0.69$$

$$Var(\ln(\widehat{OR})) = \frac{1}{20} + \frac{1}{189} + \frac{1}{392} + \frac{1}{7372}$$
$$= 0.058$$

$$IC_{95\%}(\ln(OR)) \coloneqq 0.69 \pm 1{,}96\sqrt{0.058}$$
$$[\ 0.218\,;\, 1.162\ ]$$

$$IC_{95\%}(OR) \coloneqq \quad [\ e^{0.218}\,;\, e^{1.162}\ ]$$
$$[\ 1.24\,;\ 3.20\ ]$$

Test d'une association : test du Chi-2

	E+	E-	
M+	a	b	m_1
M-	c	d	m_0
	n_1	n_0	n

Test de l'hypothèse H_0 contre H_1

$$H_0 : \; P(M+ \mid E+) \; = \; P(M+ \mid E-)$$

$$\Leftrightarrow H_0 : \; RR = 1 \quad (ou \quad OR = 1)$$

$$H_1 : \; P(M+ \mid E+) \; \neq \; P(M+ \mid E-)$$

$$\Leftrightarrow H_1 : \; RR \neq 1 \quad (ou \quad OR \neq 1)$$

- Test d'indépendance du Chi-2 : sous H_0 , $C_{ij} = \frac{n_i m_j}{n}$

$$\mathrm{X}_a = \sum_{\substack{i=0,1 \\ j=0,1}} \frac{(O_{ij} - C_{ij})^2}{C_{ij}} = \frac{n(ad - bc)^2}{n_1 n_0 m_1 m_0} \quad et \quad \mathrm{X}_a \equiv \chi^2(1)$$

$$\mathrm{X}_b = \frac{\left[a - \frac{n_1 m_1}{n}\right]^2}{\frac{n_1 n_0 m_1 m_0}{n^2(n-1)}} = \frac{n}{(n-1)} \mathrm{X}_a \quad et \quad \mathrm{X}_b \equiv \chi^2(1)$$

- Conditions d'application : $C_{ij} \geq 5$ *pour tout i et j*.

Test d'une association : test du Chi-2

- Au niveau α, $H_0 : \; RR = 1 \quad (ou\; OR = 1)$
  - Rejet de H_0 si $\mathrm{X}_a \geq \chi^2{}_\alpha$ où $\chi^2{}_\alpha$ est le quantile de niveau α d'une $\chi^2(1)$
  - $p = P(\chi^2(1) > \mathrm{X}_a)$ *p* degrés de signification
  - Rq : Rejet de H_0 si l'IC du RR ou de l'OR ne contient pas la valeur 1

- Test unilatéral $H_0 : \; RR = 1 \quad contre \quad H_1 \; : \; RR > 1$
  - Rejet de $H_0 \; si \; \mathrm{X}_a \geq \chi^2{}_{2\alpha}$ et si $\widehat{RR} > 1$
  - $p = \frac{1}{2} P(\chi^2(1) > \mathrm{X}_a)$ *p* degrés de signification

Test d'une association : test du Chi-2

- Si $3 \leq C_{ij} < 5$: Test du Chi-2 corrigé de Yates

$$\mathrm{X}_c = \sum_{\substack{i=0,1 \\ j=0,1}} \frac{\left(\left|O_{ij} - C_{ij}\right| - 0{,}5\right)^2}{C_{ij}} = \frac{n\left(|ad - bc| - \frac{n}{2}\right)^2}{n_1 n_0 m_1 m_0} \quad et \quad \mathrm{X}_c \equiv \chi^2(1)$$

- Si $C_{ij} < 3$: Test exact de Fisher

↳ Consiste à calculer les p-values exactes en calculant les probabilités d'obtenir chacun des tableaux possibles

Exemple 1 (suite)

		Alcool (consommation quotidienne) ≥ 40 cl	<40 cl	
Hypotrophie fœtale	M+	20	189	209
	M-	392	7372	7764
		412	7561	7973

- Test de l'association

$$X_a = \frac{7973 \times (20 \times 7372 - 189 \times 392)^2}{412 \times 7561 \times 209 \times 7764} = 8.49$$

$$8.49 = X_a > {\chi^2}_{0.05} = 3.84 \longrightarrow \text{Rejet de } H_0$$

$$P(\chi^2(1) > 8.49) = p < 0.01$$

- Mesures d'association

$\widehat{RR} = 1.94$ $\quad IC_{95\%}(RR) : [\,1.25\,;\,3.07\,]$

$\widehat{OR} = 1.99$ $\quad IC_{95\%}(OR) : [\,1.24\,;\,3.20\,]$

Mesures d'association et type d'enquête

		Exposition		
		E+	E-	
Maladie	M+	a	b	m_1
	M-	c	d	m_0
		n_1	n_0	n

$$\hat{P}_1 = \hat{P}(M + | E +) = \frac{a}{n_1} \qquad \hat{P}_0 = \hat{P}(M + | E -) = \frac{b}{n_0}$$

$$\hat{P}_{E1} = \hat{P}(E + | M +) = \frac{a}{m_1} \qquad \hat{P}_{E0} = \hat{P}(E + | M -) = \frac{c}{m_0}$$

- Risque relatif :

$$\widehat{RR} = \frac{\hat{P}_1}{\hat{P}_0} = \frac{a/n_1}{b/n_0} \neq \frac{\hat{P}_{E1}}{\hat{P}_{E0}}$$

- Odds ratio :

$$\widehat{OR} = \frac{\hat{P}_1/(1-\hat{P}_1)}{\hat{P}_0/(1-\hat{P}_0)} = \frac{ad}{bc} = \frac{\hat{P}_{E1}/(1-\hat{P}_{E1})}{\hat{P}_{E0}/(1-\hat{P}_{E0})}$$

Enquête transversale

		Exposition		
		E+	E-	
Maladie	M+	a	b	m_1
	M-	c	d	m_0
		n_1	n_0	n

- Prévalence de M+ chez les E+ et les E- estimables : $\hat{P}_1 = \frac{a}{n_1}$ et $\hat{P}_0 = \frac{b}{n_0}$
- Prévalence de E+ chez les M+ et les M- estimables : $\hat{P}_{E1} = \frac{a}{m_1}$ et $\hat{P}_{E0} = \frac{c}{m_0}$
- Risque relatif et odds ratio sont estimables

$$\widehat{RR} = \frac{\hat{P}_1}{\hat{P}_0} = \frac{a/n_1}{b/n_0} \qquad \widehat{OR} = \frac{\hat{P}_1/(1-\hat{P}_1)}{\hat{P}_0/(1-\hat{P}_0)} = \frac{\hat{P}_{E1}/(1-\hat{P}_{E1})}{\hat{P}_{E0}/(1-\hat{P}_{E0})} = \frac{ad}{bc}$$

Enquête exposé – non exposé

		Exposition		
		E+	E-	
Maladie	M+	a	b	m_1
	M-	c	d	m_0
		n_1	n_0	n

$$\underbrace{n_1 \qquad n_0}_{\textbf{Fixes}}$$

- Prévalence de M+ chez les E+ et les E- estimables : $\hat{P}_1 = \frac{a}{n_1}$ et $\hat{P}_0 = \frac{b}{n_0}$
- Prévalence de E+ chez les M+ et les M- **non** estimables
- Risque relatif et odds ratio sont estimables

$$\widehat{RR} = \frac{\hat{P}_1}{\hat{P}_0} = \frac{a/n_1}{b/n_0} \qquad \widehat{OR} = \frac{\hat{P}_1/(1-\hat{P}_1)}{\hat{P}_0/(1-\hat{P}_0)} = \frac{ad}{bc}$$

Enquête cas-témoins

		Exposition			
		E+	E-		
Maladie	M+	a	b	m_1	Fixes
	M-	c	d	m_0	
		n_1	n_0	n	

- Prévalence de M+ chez les E+ et les E- **non** estimables
- Prévalence de E+ chez les M+ et les M- estimables : $\hat{P}_{E1} = \frac{a}{\mathrm{m}_1}$ et $\hat{P}_{E0} = \frac{c}{\mathrm{m}_0}$
- Risque relatif **non** estimable
- Odds ratio estimable par l'intermédiaire de $\hat{P}_{E1}$ et $\hat{P}_{E0}$

$$\widehat{OR} = \frac{\hat{P}_{E1} / (1 - \hat{P}_{E1})}{\hat{P}_{E0} / (1 - \hat{P}_{E0})} = \frac{ad}{bc}$$

Choix des mesures d'association

- Risque relatif
    - Interprétation simple
    - Non estimable dans les enquêtes cas-témoins
- Odds ratio
    - Estimable dans tous les types d'enquêtes
    - Interprétation : $OR \approx RR$ quand la maladie est rare dans la population
- Relation entre risque relatif et odds ratio

$$\boxed{RR = \frac{OR}{1 + P_0(OR - 1)} \quad et \quad OR = \frac{RR(1 - P_0)}{1 - P_0 \times RR}}$$

Si $P_0 = P(M+|E-)$ et $P_1 = P(M+|E+) = P_0 \times RR$ petits alors $OR \approx RR$

II. Notion de causalité

- Enquêtes expérimentales randomisées
  - difficultés pratiques

↳ Preuve de causalité

- Enquêtes observationnelles
  - Enquête transversale, cas-témoins, cohorte
  - Souvent les seules possibles
  - Pas d'interprétation causale

↳ Présomption de causalité

Critères de présomption causale de Bradford Hill

Critères internes à l'étude

1. Force de l'association
2. Relation dose- effet
3. Pas d'ambiguïté sur la chronologie
4. Spécificité de l'association

Critères internes à l'étude

1. Force de l'association « strength »

Type de cancer	Facteur de risque	Risque Relatif
Sein	Radiations > 100 rad	3
Poumon	Cigarette > 25 cig/ j	10
Oesophage	Alcool > 100g/ j	17.5
Mésothéliome	Amiante (exp. prof.)	>200
Vessie	Benzidine (exp. Prof.)	500

2. Relation dose- effet « biologic gradient »

Relation tabac – cancer de la vessie

Nb de cigarettes fumées par jour	Risque Relatif
Non fumeur	1
1 - 19 cig/ j	2.9
20 - 39 cig/ j	3.5
>= 40 cig/ j	4

Critères internes à l'étude

3. Chronologie « temporality » : l'exposition (la cause) doit précéder l'effet

- Problème si enquête transversale ou cas-témoins
- Ex : Lien entre cancer du pancréas et diabète

		Diabète	
		oui	non
Cancer pancréas	oui	a	b
	non	c	d

?

Diabète $\rightleftarrows$ Cancer pancréas

4. Spécificité des associations : la cause doit conduire à un seul effet

- Critère critiqué
- Par exemple: Tabac associé au cancer du poumon et aux pathologies cardio-vasculaires
- Ex : Exposition au Benzène associée à un type bien particulier de leucémie

Critères de présomption causale de Bradford Hill

Critères externes à l’étude

5. Constance des résultats dans diverses études (reproductibilité)
6. Plausibilité biologique de l'hypothèse
7. Cohérence des résultats avec les hypothèses qui ont conduit à la mise en œuvre de l’étude (cohérence interne des résultats d’un sous-groupe à l’autre)

III. Mesures d'impact potentiel

- Exemple : Amiante est la cause de la maladie

- Quelle est la proportion de cas attribuables au l'amiante ?

- Nombre de malade : $M = P(M+) \times N$

Prévalence de la maladie

taille de l'échantillon

- Nombre de malades à cause de l'exposition :

$$M^* = M - P(M + \mid E-) \times N$$

prévalence de la maladie
chez les non exposés E-

Risque attribuable à l’exposition

- $P_M = P(M+)$ $\qquad P_0 = P(M+ \mid E-)$ $\qquad P_1 = P(M+ \mid E+)$

  $P_E = P(E+)$ $\qquad RR = \frac{P_1}{P_0}$

$$P_M = P_E P_1 + (1 - P_E) P_0 = P_0 P_E \left(\frac{P_1}{P_0} - 1 \right) + P_0$$

- Risque attribuable à une exposition

$$\boxed{RA = \frac{M^*}{M} = \frac{P_M - P_0}{P_M} = \frac{P_E(RR - 1)}{P_E(RR - 1) + 1}}$$

Exemple

- Risque relatif entre une maladie et l'amiante $RR = 9.77$

- Prévalence de l'amiante dans la population $P_E = 5\%$

$$\widehat{RA} = \frac{\widehat{P_E}\left(\widehat{RR} - 1\right)}{\widehat{P_E}\left(\widehat{RR} - 1\right) + 1} = \frac{0.05(9.77 - 1)}{0.05(9.77 - 1) + 1} = 0.30$$

30% des cas de la maladie sont attribuables à l'exposition à l'amiante

- Prévalence de l'amiante dans la population $P_E = 10\%$

$$\widehat{RA} = \frac{\widehat{P_E}\left(\widehat{RR} - 1\right)}{\widehat{P_E}\left(\widehat{RR} - 1\right) + 1} = \frac{0.1(9.77 - 1)}{0.1(9.77 - 1) + 1} = 0.47$$

47% des cas de la maladie sont attribuables à l'exposition à l'amiante

Risque attribuable et type d'enquête

Enquête transversale et enquête de cohorte où P_E est estimable

$$\widehat{RA} = \frac{\widehat{P_E}(\widehat{RR} - 1)}{\widehat{P_E}(\widehat{RR} - 1) + 1}$$

Enquête cas-témoins

RR, P_M et P_E non estimables

$$P_E = P_M P(E + |M+) + (1 - P_M) P(E + |M-)$$

Si la maladie est rare dans la population,

$$\widehat{RA} = \frac{\widehat{P_{E0}}(\widehat{OR} - 1)}{\widehat{P_{E0}}(\widehat{OR} - 1) + 1}$$

Prévalence de l'exposition
chez les témoins $\hat{P}(E + |M-)$

Intervalles de confiance

	E+	E-	
M+	a	b	m_1
M-	c	d	m_0
	n_1	n_0	n

$$\widehat{RA} = \frac{\widehat{P_E}\left(\widehat{RR}-1\right)}{\widehat{P_E}\left(\widehat{RR}-1\right)+1}$$

- $\ln(1-\widehat{RA})$ converge vers une loi normale quand $n \to \infty$

$$IC_\alpha(\ln(1-RA)) := [B_{inf}; B_{sup}] \quad = \quad \ln(1-\widehat{RA}) \pm z_{\frac{\alpha}{2}}\sqrt{Var\left(\ln(1-\widehat{RA})\right)}$$

- Intervalle de confiance de niveau α pour RA

$$IC_\alpha(RA) := \left[1-e^{B_{sup}}; 1-e^{B_{inf}}\right]$$

Intervalle de confiance

$$IC_\alpha(\ln(1 - RA)) \coloneqq \ln(1 - \widehat{RA}) \pm z_{\frac{\alpha}{2}} \sqrt{Var\left(\ln(1 - \widehat{RA})\right)}$$

- Enquête transversale

$$\boxed{Var\left(\ln(1 - \widehat{RA})\right) = \frac{1}{n}\left(\frac{c + \widehat{RA}(a + d)}{b}\right)}$$

- Enquête cas-témoins

$$\boxed{Var\left(\ln(1 - \widehat{RA})\right) = \frac{a}{b \times m_1} + \frac{c}{d \times m_0}}$$

- Enquête de cohorte où P_E estimable

$$\boxed{Var\left(\ln(1 - \widehat{RA})\right) = {P_E}^2 \left(1 - \widehat{RA}\right)^4 \widehat{RR}^2 \left(\frac{c}{an_1} + \frac{c}{bn_0}\right)}$$

Exemple : enquête transversale

	E+	E-	
M+	15	30	15
M-	135	820	955
	150	850	1000

$$\widehat{P_E} = \frac{150}{1000} = 0.15$$

$$\widehat{RR} = \frac{15/150}{30/850} = 2.83$$

$$\widehat{RA} = \frac{0.15(2.83-1)}{0.15(2.83-1)+1} = 0.22$$

- $Var\left(\ln(1-\widehat{RA})\right) = \frac{1}{1000}\left(\frac{135+0.22\,(15+820)}{30}\right) = 0.0105$
- $IC_{95\%}(\ln(1-RA)) := \quad \ln(1-0.22) \pm 1.96\sqrt{0.0105} = [-0.444\,;-0.042]$
- $IC_{95\%}(RA) := [1-e^{-0.042}; 1-e^{-0.444}] = [0.04\,;0.36]$

Exemple : enquête cas-témoins

	E+	E-	
M+	70	130	200
M-	22	178	200
	92	308	400

$$\widehat{P_{E0}} = \frac{22}{200} = 0.11$$

$$\widehat{OR} = \frac{70 \times 178}{22 \times 130} = 4.36$$

$$\widehat{RA} = \frac{0.11(4.36-1)}{0.11(4.36-1)+1} = 0.27$$

- $Var\left(\ln(1-\widehat{RA})\right) = \frac{70}{130 \times 200} + \frac{22}{178 \times 200} = 0.0033$
- $IC_{95\%}(\ln(1-RA)) := \ln(1-0.27) \pm 1.96\sqrt{0.0033} = [-0.427\,;-0.202]$
- $IC_{95\%}(RA) := [1-e^{-0.202}; 1-e^{-0.427}] = [0.18\,;0.35]$