
Statistiques pour l'épidémiologie

Dr Muriel Rabilloud

Épidémiologie

- Étude
 - De la fréquence des maladies et de leurs conséquences dans les populations humaines (descriptive)
 - De leurs déterminants (analytique, étiologique)
- Observationnelle (≠ expérimentale)

2

L'épidémiologie descriptive a pour objectif de décrire la fréquence des maladies dans différentes populations et au cours du temps. Il s'agit par exemple de décrire l'hétérogénéité de la fréquence des cancers selon la zone géographique ou l'évolution de la fréquence des cancers avec le temps. Ces études vont permettre de soulever des hypothèses sur les facteurs associés au risque de cancer.

Les études d'épidémiologie analytique ont pour objectif de quantifier l'association entre des facteurs d'exposition et le risque de développer une maladie donnée. Il s'agit par exemple de quantifier le lien entre l'exposition à l'amiante et le risque de cancer du poumon.

Contrairement aux études expérimentales dont l'exemple le plus classique est l'essai clinique randomisé, dans les études observationnelles il n'y a pas d'intervention.

On observe par exemple des sujets qui sont exposés ou non à un facteur que l'on pense être associé à une augmentation du risque de développer une maladie.

L'investigateur ne contrôle pas l'exposition au facteur de risque.

Plan du cours

- Epidémiologie descriptive
 - Prévalence d'une maladie
 - Incidence d'une maladie

- Epidémiologie analytique
 - Etude de cohortes exposés non exposés
 - Etude cas-témoins

Épidémiologie descriptive

Épidémiologie descriptive

- Fournit des indicateurs de santé = Mesurer l'importance d'un événement dans une population donnée
- Proportion de cas à un moment donné
→ **Prévalence**
- Mesure de la vitesse d'apparition des nouveaux cas
→ **Taux d'Incidence**

Types d'études permettant de mesurer l'importance d'un événement

Les études transversales (cross sectional studies)



Prévalence de la maladie

Les études de cohorte (cohort studies)



Incidence de la maladie

6

Si l'on se place sur une échelle de temps, une étude transversale consiste à faire une coupe à un moment donné et à mesurer la fréquence de la maladie sur cette coupe. Il s'agit d'une photographie de la population à un moment donné.

Les études qui vont permettre de mesurer la vitesse d'apparition de la maladie dans une population sont les études de cohorte. La différence principale entre une étude transversale et une étude de cohorte porte sur le suivi des sujets qui vont participer à l'étude.

Dans une étude transversale, il n'y a pas de suivi des sujets. On les observe à un moment donné et ils ont ou ils n'ont pas la maladie.

Dans une étude de cohorte, que l'on appelle également étude longitudinale, les sujets n'ont pas la maladie étudiée au moment de leur entrée dans la cohorte. Ils vont être suivis un temps plus ou moins long et on recueille l'apparition ou non de la maladie au cours du suivi.

Un exemple d'études transversales

- Etudes transversales réalisées un jour donné dans tous les hôpitaux en France
 - Permet d'avoir une mesure de la prévalence des infections nosocomiales
 - Ces études sont répétées tous les 5 ans environ
 - L'évolution de la prévalence est un indicateur de l'efficacité de la lutte contre les infections nosocomiales
-

Un exemple célèbre d'étude de cohorte : la cohorte de Framingham

- Suivi de 5209 habitants de la ville de Framingham (Massachusetts, Etats Unis),
- Début de l'étude en 1948,
- Etude de l'apparition des évènements cardio-vasculaires,
- A contribuer à la connaissance des facteurs de risque des maladies cardio-vasculaires

<http://www.framinghamheartstudy.org/>

8

La première cohorte de Framingham était constituée de 5209 hommes et femmes âgés de 30 à 62 ans qui n'avaient pas d'antécédent d'infarctus du myocarde ou d'accident vasculaire cérébral et pas de symptômes de maladie cardiovasculaire au moment de l'inclusion.

Entre 1950 et 2010, plus de 2000 articles scientifiques reposant sur les données de l'étude Framingham ont été publiés.

Prévalence d'une maladie

Proportion de malades présents dans une population à un moment donné :

$$P = \frac{M}{N}$$

← effectif de malades*
← effectif total de sujets**

* Par exemple : diabète insulino-dépendant, insuffisance ventriculaire gauche ou maladie de durée limitée, grippe hivernale,...

** Par exemple : population française de 20 ans ou plus

La prévalence est fonction de

- ▶ De l'incidence (vitesse d'apparition des nouveaux cas)
- ▶ De la durée de la maladie



Prévalence au 1 juin 2001 = 4 /population Prévalence au 1 juin 2004 = 2 /population

Durée de maladie courte : prévalence faible

Prenons l'exemple théorique de 2 maladies qui ont la même incidence (même vitesse d'apparition des nouveaux cas), mais des durées moyennes différentes. La maladie dont la durée moyenne est plus grande a une prévalence plus élevée.

Prévalence d'une maladie

- Mesure indirecte de l'incidence
- Permet d'évaluer les besoins de prise en charge médicale

Estimation d'une prévalence

- Constitution d'un échantillon représentatif de la population
- Utilisation d'un estimateur = la proportion observée sur l'échantillon notée

$$\hat{P}$$

12

Un échantillon est représentatif d'une population si il ressemble à la population. Le caractère représentatif ou non d'un échantillon dépend de la façon dont a été constitué l'échantillon.

Un estimateur correspond à une fonction utilisée pour calculer dans notre cas la prévalence.

Le chapeau sur P permet d'indiquer que la valeur calculée sur l'échantillon est une estimation de la prévalence dans la population.

Exemple d'estimation de prévalence (1)

Prévalence des infections nosocomiales dans les hôpitaux français en 1996

Etude transversale un jour donné

Echantillon de 46 000 patients hospitalisés

3370 patients avec une infection nosocomiale

$$\hat{p} = \frac{3370}{46000} = 7,3\%$$

13

On suppose que l'échantillon constitué un jour donné est représentatif de la population des patients hospitalisés au cours de l'année 1996.

On se sert de la valeur calculée sur l'échantillon de l'étude pour parler de la prévalence des infections nosocomiales dans les hôpitaux français en 1996.

Il s'agit d'un exemple d'inférence statistique.

Le résultat communiqué comprend l'estimation ponctuelle (la prévalence en 1996 est estimée à 7,3%), et l'intervalle de confiance à 95% qui donne la précision de l'estimation.

En 2006, la prévalence était estimée à 4%.

Exemple d'estimation de prévalence (2)

- Approximation de la distribution de la prévalence par une distribution normale :

$$N\left(0,073, \sqrt{\frac{0,073 \times (1-0,073)}{46000}}\right)$$

- Intervalle de confiance à 95% :

$$0,073 \pm 1,96 \times \sqrt{\frac{0,073 \times (1 - 0,073)}{46000}}$$

Exemple d'estimation de prévalence (3)

- Intervalle de confiance à 95% : $0,073 \pm 0,002$
- La précision de l'estimation est d'environ 0,2%
- IC à 95% : [7,1% - 7,5%]
- Conditions d'approximation par la distribution normale vérifiées : $n \geq 30$, $nf_1 \geq 5$, $nf_2 \geq 5$, $n(1-f_1) \geq 5$, $n(1-f_2) \geq 5$

Incidence d'une maladie

- Taux d'incidence mesure la « vitesse d'apparition » des nouveaux cas

- Taux d'incidence mesure la probabilité pour un individu de développer la maladie par unité de temps

Taux d'incidence d'une maladie

En théorie

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt | T > t)}{dt}$$

T = temps écoulé jusqu'à la survenue de l'événement étudié

dt = longueur de l'intervalle de temps

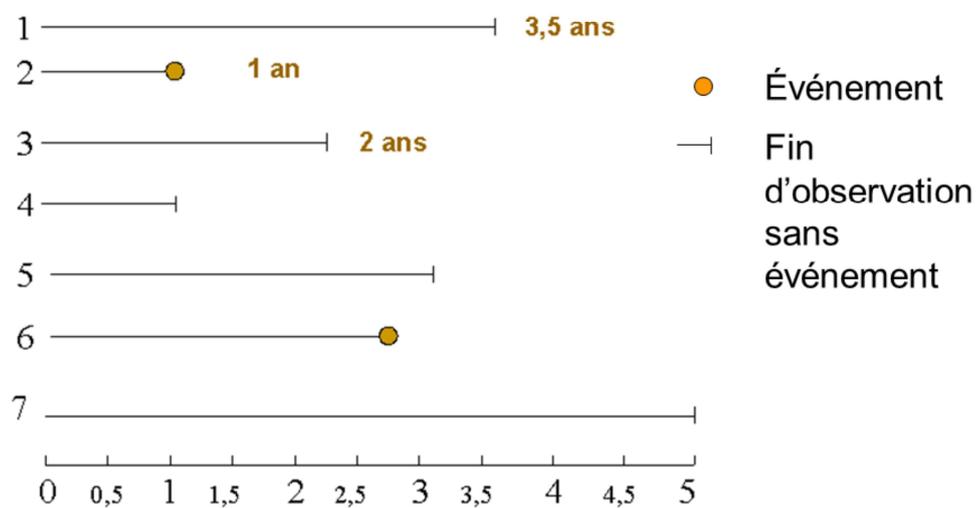
Dimension dynamique des taux d'incidence et de mortalité

17

Le taux d'incidence n'est pas une probabilité, mais une probabilité par unité de temps.

Il s'agit formellement du rapport d'une probabilité conditionnelle (la probabilité que l'événement survienne dans l'intervalle de temps dt sachant qu'il ne s'est pas produit avant t) sur la longueur de l'intervalle de temps.

Estimation du taux d'incidence d'une maladie



Estimation du taux d'incidence d'une maladie (2)

- Deux événements au cours du suivi
- Mesure du temps en année
- Longueur de l'intervalle de temps de 0,5 année
- Nombre d'intervalle de temps de longueur 0,5 est de 36

$$\hat{\lambda} = \frac{2/36}{0,5} = \frac{0,055}{0,5} = 0,11 \text{ par personne année}$$

19

La probabilité pour une personne de développer la maladie sur un intervalle de temps de ½ année est de 0,055 (5,5%).

Elle est de 0,11 pour une année.

On dit que le taux d'incidence est estimé à 0,11 par personne année.

Il mesure bien la vitesse d'apparition de la maladie.

Lorsque l'évènement étudié est le décès, on parle de force de mortalité.

Estimation du taux d'incidence d'une maladie (3)

- Basée sur l'observation d'un groupe d'individus



Hypothèse = tous les individus ont le même taux

- Basée sur l'observation d'une période de suivi



Hypothèse = le taux est constant sur la période

Estimation du taux d'incidence d'une maladie (4)

- Sous ces hypothèses les individus et les intervalles de temps sont interchangeables
- Suivre 10 personnes pendant 1 demi-année est équivalent à suivre 5 personnes pendant 2 demi-années = 5 personnes années
- Dans notre exemple le nombre de personnes années est de 18 (36 intervalles de temps de longueur 0,5 année)

Estimation du taux d'incidence d'une maladie (5)

$$\hat{\lambda} = \frac{\text{nombre de nouveaux cas}}{\text{nombre de personnes année}} = \frac{2}{18}$$

= 0,11 par personne année

L'estimation du taux d'incidence de la maladie est
de 11 nouveaux cas pour 100 personnes année

Exemple d'études de cohorte visant à estimer le taux d'incidence d'une maladie (1)

- Etude de cohorte pour estimer le taux d'incidence de la cirrhose décompensée dans une population à risque
- Observation de 56 nouveaux cas de cirrhose décompensée
- Observation de 40 000 personnes-année

**Exemple d'études de cohorte visant à estimer
le taux d'incidence d'une maladie (2)**

$$\hat{\lambda} = \frac{56}{40\ 000} = 0,0014$$

= 14 pour 10 000 personnes année

Exemple d'études de cohorte visant à estimer le taux d'incidence d'une maladie (3)

- Événement rare et nombre de personnes année élevé
- La distribution du nombre de nouveaux cas de cirrhose M suit une distribution de Poisson de paramètre $M = 56$
- La variance du nombre de nouveaux cas de cirrhose = $M = 56$

Exemple d'études de cohorte visant à estimer le taux d'incidence d'une maladie (4)

$$\begin{aligned}\text{variance}(\hat{\lambda}) &= \frac{\text{variance}(M)}{\text{Nombre de personnes année}^2} \\ &= \frac{M}{\text{Nombre de personnes année}^2} \\ &= \frac{\hat{\lambda}}{\text{Nombre de personnes année}} \\ &= \frac{0,0014}{40000}\end{aligned}$$

Exemple d'études de cohorte visant à estimer le taux d'incidence d'une maladie (4)

- $M > 50$, distribution du taux d'incidence approximée par une distribution normale :

$$N\left(0,0014, \sqrt{\frac{0,0014}{40000}}\right)$$

- Intervalle de confiance à 95% :

$$0,0014 \pm 1,96 \times \sqrt{\frac{0,0014}{40000}}$$

Exemple d'études de cohorte visant à estimer le taux d'incidence d'une maladie (5)

- Intervalle de confiance à 95% :
 $0,0014 \pm 0,0004$
- La précision de l'estimation est d'environ 4 pour 10000 personnes années
- L'intervalle de confiance à 95% du taux d'incidence est compris entre 10 et 18 pour 10000 personnes année

Risque de maladie (1)

- Incidence cumulée de la maladie jusqu'à un temps t

$$R(t) = P(T \leq t)$$

- Même durée de suivi pour tous les individus qui n'ont pas développé la maladie

$$\hat{R}(t) = \frac{M}{N}$$

M = nombre de nouveaux cas sur la période de suivi

N = nombre de sujets à risque au début de la période

Le risque correspond à une probabilité. Il s'agit de la probabilité d'avoir développé la maladie au temps t .

Exemple d'estimation du risque de récurrence d'une maladie

- Cohorte de 368 femmes opérées d'un cancer du sein et suivies pendant 5 ans
- Estimation du risque de récurrence à 5 ans

$$\hat{r}_{5 \text{ ans}} = \frac{130}{368} \approx 35\%$$

30

Tous les sujets ont eu la maladie.

L'événement étudié est la récurrence de la maladie.

Risque de maladie (2)

- Durée de suivi variable d'un individu à l'autre
- Taux d'incidence constant sur la période considérée

$$R(t) = 1 - e^{-\lambda t}$$

λ = taux d'incidence

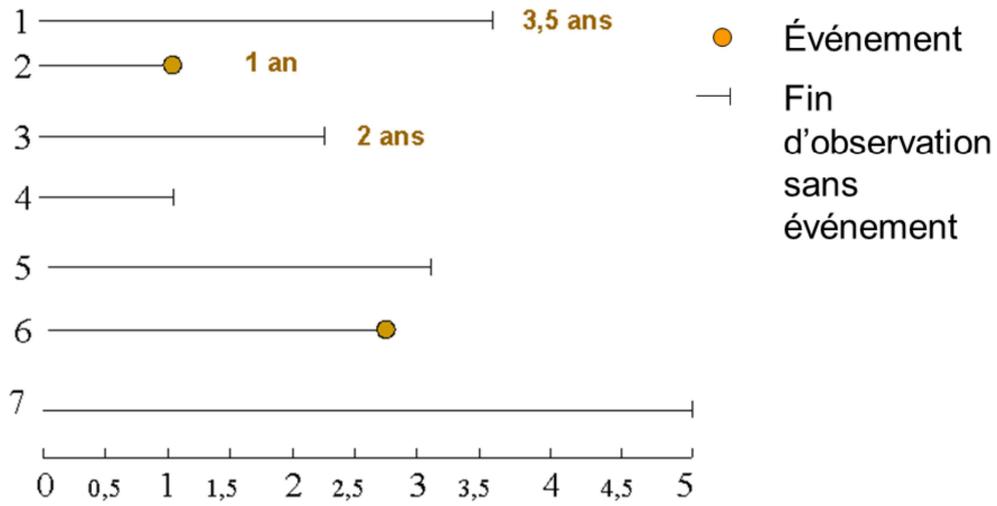
t = période de suivi

31

La probabilité de développer la maladie avant le temps t est égale à 1 moins la probabilité de ne pas l'avoir développé.

$\exp(-\lambda t)$ = probabilité de ne pas avoir développé la maladie au temps t
 λt correspond au taux cumulé .

Estimation du risque de maladie sur une période de 3 ans



Estimation du risque de maladie sur une période de 3 ans (suite)

$$\hat{R}(t) = 1 - e^{-\hat{\lambda}t}$$

$$\begin{aligned}\hat{r}_{3\text{ans}} &= 1 - e^{-0,11 \times 3} \\ &= 0,28\end{aligned}$$

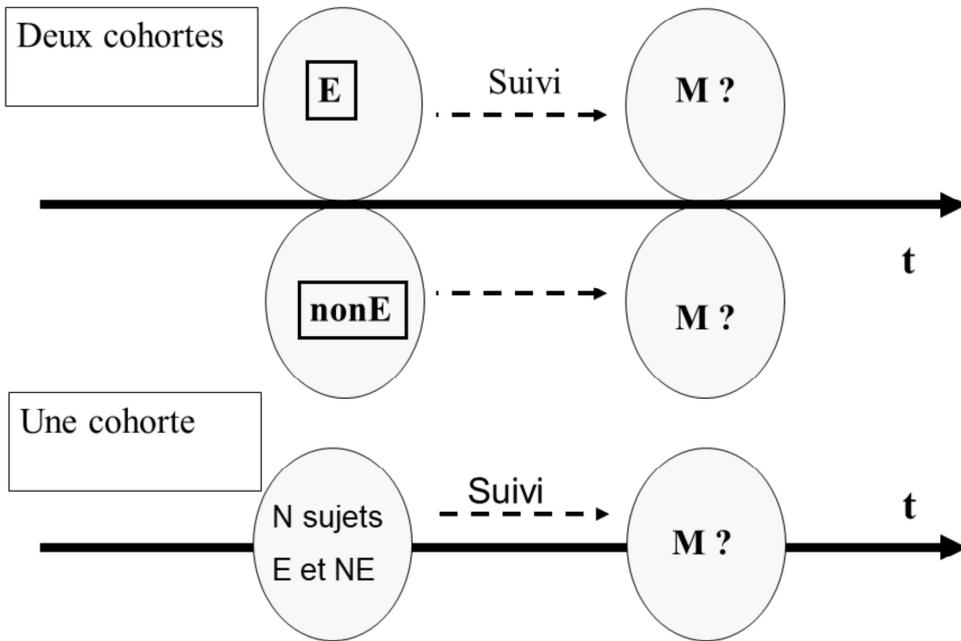
Le risque est estimé à 28% à 3 ans

Épidémiologie analytique

Etudes d'observation

- Etudes de cohorte
 - ❖ Identification de facteurs de risque de survenue de la maladie
 - ❖ Identification de facteurs pronostiques
- Etudes cas-témoins
 - ❖ Identification de facteurs de risque de survenue de la maladie

$$P [M / E] > P [M / \text{non E}] \quad ?$$



36

Étude de cohorte

- Cohorte de 368 femmes opérées pour un cancer du sein et suivies pendant 5 ans

- Objectif de l'étude :

Évaluer la valeur pronostique du type cytologique de la tumeur sur le risque de récurrence à 5 ans

- Critère de résultat :

Récurrence à 5 ans oui ou non

37

Le facteur étudié est le type cytologique de la tumeur.

Estimation du risque de récurrence à 5 ans selon le type cytologique de la tumeur

- Type cytologique 3 :

$$\hat{r}_{\text{Type 3}} = \frac{96}{201} \approx 48\%$$

- Type cytologique 1 ou 2 :

$$\hat{r}_{\text{Type 1 ou 2}} = \frac{34}{167} \approx 20\%$$

38

L'hypothèse à tester dans cette étude est que le risque de récurrence des femmes qui avaient une tumeur de type cytologique 3 est différent du risque de récurrence des femmes qui avaient une tumeur de type 1 ou 2.

Étude de cohorte

Questions :

- Existe-t-il une relation entre le type cytologique et la récurrence à 5 ans?
- Quelle est la force de l'association entre le type cytologique et la récurrence à 5 ans?

Étude de cohorte

Tableau de contingence

	Récidive	Pas Récidive	
Type 3	96	105	201
Type 1 ou 2	34	133	167
	130	238	368

Étude de cohorte

Existe-t-il une association? → Test du chi²

	Récidive	Pas Récidive	
Type 3	96 (71)	105 (130)	201
Type 1 ou 2	34 (59)	133 (108)	167
	130	238	368

Étude de cohorte

$$\begin{aligned}\chi_1^2 &= \frac{(96-71)^2}{71} + \frac{(105-130)^2}{130} \\ &+ \frac{(34-59)^2}{59} + \frac{(133-108)^2}{108} \\ &= 29,99\end{aligned}$$

Conclusion

$\chi^2 > 3,84$ donc la probabilité que la différence observée soit due au hasard est inférieure à 5%

La différence est statistiquement significative

Le risque de récurrence à 5 ans est plus élevé chez les femmes ayant un type cytologique 3 que chez celles ayant un type cytologique 1 ou 2, et cette différence est statistiquement significative

Mesure de la force de l'association par le risque relatif

Quelle est la force de cette association?

→ Risque relatif

	Récidive	Pas Récidive	
Type 3	a	b	m_0
Type 1 ou 2	c	d	m_1
	n_0	n_1	n

Risque relatif

Risque Relatif = Ratio de Risques

$$RR = \frac{\text{Risque de récurrence}_{\text{Type 3}}}{\text{Risque de récurrence}_{\text{Type 1 ou 2}}}$$

Estimation du risque relatif (1)

$$\hat{RR} = \frac{a/m_0}{c/m_1} = \frac{0,48}{0,20} = 2,4$$

Les femmes qui ont une tumeur de type 3 ont 2,4 fois plus de risque de récidiver à 5 ans que les femmes ayant une tumeur de type 1 ou 2

Estimation du risque relatif (2)

- Le risque relatif de récurrence chez les femmes ayant une tumeur de type 3 par rapport aux femmes ayant une tumeur de type 1 ou 2 est estimé à 2,4
- L'intervalle de confiance à 95% du risque relatif estimé est compris entre 1,8 et 3,3

Risque Relatif et Taux Relatif

$$RR = \frac{R_1(t)}{R_0(t)} \quad TR(t) = \frac{\lambda_1(t)}{\lambda_0(t)}$$

Si le taux relatif est constant au cours de la période de suivi et que l'événement est rare ou la durée d'observation courte

$$TR \approx \frac{R_1(t)}{R_0(t)} = RR$$

Dans ce cas le risque relatif peut être interprété comme un taux relatif.

Taux d'incidence de la récurrence (1)

- Taux d'incidence de la récurrence dans le groupe de femmes ayant un type cytologique 3 :

96

Nbr personnes année

- Taux d'incidence de la récurrence dans le groupe de femmes ayant un type cytologique 1 ou 2 :

34

Nbr de personnes année

Taux d'incidence de la récidive (2)

- Nombre de personnes année de suivi :
 - 5 personnes année pour les femmes qui n'ont pas récidivé dans les 5 ans
 - Délai en années entre l'intervention et la récidive pour celles qui ont récidivé dans les 5 ans

- Approximation en considérant que les femmes qui ont récidivé, ont été suivies en moyenne 2,5 ans

Taux d'incidence de la récurrence (3)

- Taux d'incidence de la récurrence dans le groupe cytologique 3 :

$$\frac{96}{(105 \times 5 + 96 \times 2,5)} = \frac{96}{765} \approx 0,125$$

- Taux d'incidence de la récurrence dans le groupe cytologique 1 ou 2 :

$$\frac{34}{(133 \times 5 + 34 \times 2,5)} = \frac{34}{750} \approx 0,045$$

Taux relatif de récidence

$$\text{Taux relatif} = \frac{0,125}{0,045} \approx 2,8$$

Le taux relatif de récidence est estimé à 2,8

Le risque relatif de récidence à 5 ans est
estimé à 2,4

Mesure de la force de l'association par l'odds ratio

Odds Ratio = Ratio d'Odds

$$\text{Odds de Récidive}_{\text{Type 3}} = \frac{R_{\text{Type 3}}}{1 - R_{\text{Type 3}}}$$

$$\text{Odds de Récidive}_{\text{Type 1 ou 2}} = \frac{R_{\text{Type 1 ou 2}}}{1 - R_{\text{Type 1 ou 2}}}$$

Mesure de la force de l'association par l'odds ratio

Quelle est la force de cette association?

→ odds ratio

	Récidive	Pas Récidive	
Type 3	a	b	m_0
Type 1 ou 2	c	d	m_1
	n_0	n_1	n

Estimation de l'odds de récurrence dans chaque groupe

$$\hat{\text{Odds}}_{\text{Type 3}} = \frac{\frac{a}{m_0}}{1 - \frac{a}{m_0}} = \frac{\frac{a}{m_0}}{\frac{b}{m_0}} = \frac{a}{b}$$

$$\hat{\text{Odds}}_{\text{Type 1 ou 2}} = \frac{\frac{c}{m_1}}{1 - \frac{c}{m_1}} = \frac{\frac{c}{m_1}}{\frac{d}{m_1}} = \frac{c}{d}$$

Estimation de l'odds ratio de récidence (1)

$$\begin{aligned}\hat{\text{OR}} &= \frac{a/b}{c/d} = \frac{a \times d}{c \times b} \\ &= \frac{96 \times 133}{34 \times 105} = 3,6\end{aligned}$$

Les femmes ayant une tumeur de type 3 ont un odds de récidence multiplié par 3,6 par rapport à celles ayant une tumeur de type 1 ou 2

Estimation de l'odds ratio de récurrence (2)

- L'odds ratio de récurrence chez les femmes ayant une tumeur de type 3 par rapport aux femmes ayant une tumeur de type 1 ou 2 est estimé à 3,6
- L'intervalle de confiance à 95% de l'odds ratio estimé est compris entre 2,2 et 5,7

Événement rare dans la population

- La probabilité de ne pas avoir l'événement est proche de 1 :

$$1 - \text{Risque} \approx 1$$

- L'odds de l'événement est proche du risque

$$\text{Odds} = \frac{\text{Risque}}{1 - \text{Risque}} \approx \text{Risque}$$

Risque Relatif et Odds Ratio

- L'odds ratio est toujours supérieur au risque relatif
- Lorsque l'événement d'intérêt est rare dans la population étudiée
 - L'odds ratio est proche du risque relatif
 - L'odds ratio peut s'interpréter comme un risque relatif

Mesure de la force de l'association dans une étude de cohorte

- Taux relatif de l'événement
- Risque Relatif de l'événement à un temps donné
- Odds Ratio de l'événement à un temps donné