

Fiche mémo biostatistique / épidémiologie clinique

1. Notation pour la fiche mémo

- « Exposition » = toute caractéristique individuelle, intrinsèque ou extrinsèque
- « M » = « maladie » = état de santé d'un individu dont on cherche à identifier les expositions qui lui sont (causalement) associées
- « E » = exposition d'intérêt = exposition dont on cherche à quantifier l'association (causale) avec M
- « $OR^*_{E \leftrightarrow M}$ » = Odds Ratio ou Risque Relatif (*Risk Ratio* ou *Hazard Ratio*) quantifiant l'association statistique entre E et M
- « Ecart systématique » = écart dû au protocole de l'étude qui se reproduirait toujours, avec le même ordre de grandeur, si on devait refaire une infinité de fois l'étude

2. Inférer statistiquement et causalement

- Inférer statistiquement = généraliser l'association statistique entre E et M estimée (dans l'échantillon) à l'association *statistique* entre E et M dans la population cible
- Inférer causalement = généraliser l'association statistique entre E et M estimée à l'association *causale* entre E et M dans la population cible

3. Intervalle de confiance à 95% ($IC_{95\%}$) d'une estimation d'un OR^*

- $[OR^*_{inf} ; OR^*_{sup}]_{95\%}$: la valeur de l' OR^* correspondant dans la population cible a 95% de chances de se trouver entre OR^*_{inf} et OR^*_{sup} (si OR^* non biaisé par du biais d'association)
- Plus l' $IC_{95\%}$ est resserré autour de l' OR^* estimé, plus l' OR^* est estimé avec précision
- « 1 » n'est pas compris dans l'intervalle de confiance à 95% $\leftrightarrow$ l' OR^* estimé est significativement différent de « 1 » $\leftrightarrow$ association statistique significative entre E et M

4. Risques d'erreurs de 1^{ère} espèce (α) et 2^{ème} espèce (β)

- Association significative $\rightarrow$ erreur de 1^{ère} espèce = 5% (α) considérée comme faible $\rightarrow$ l'inférence est possible
- Association non significative $\rightarrow$ erreur de 2^{ème} espèce β , inconnue $\rightarrow$ l'inférence est impossible $\rightarrow$ la conclusion ne portera que sur l'échantillon (« association non significative dans l'échantillon »)

5. Puissance statistique d'une étude

- Puissance statistique = probabilité d'obtenir une différence significative dans l'échantillon lorsque qu'il existe une différence réelle $\Delta \neq 0$ dans la population cible
- Une étude a manqué de puissance statistique si elle n'a pas été capable de montrer statistiquement ($\Leftrightarrow$ association non significative) une association que l'on pense réelle
- 2 origines d'un manque de puissance statistique
 - Taille d'échantillon insuffisante
 - Association réelle entre E et M qui existe ($\Delta \neq 0$), mais elle est faible (Δ faible)

6. Analyse de survie

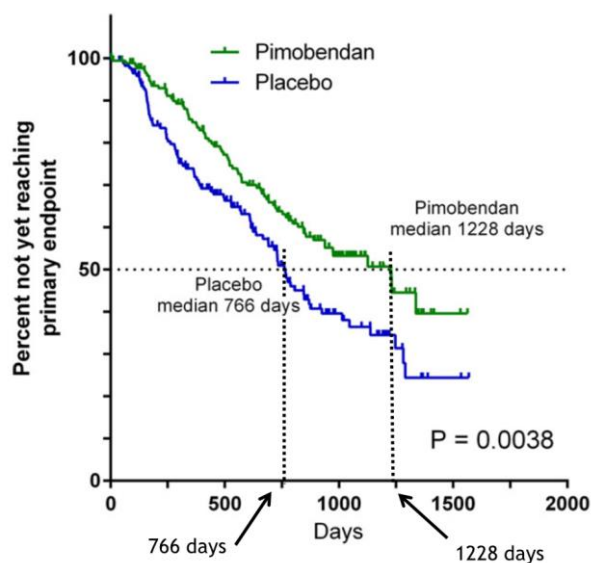
6.1. Contexte d'utilisation de l'analyse de survie

Etude clinique recrutant des individus indemnes de M, suivis au cours du temps, et dont la date de survenue de M est collectée

6.2. Définitions

- J0 = date à partir de laquelle les individus commencent à être suivis
- Événement = critère de jugement dont la survenue après J0 est étudiée
- Censure = non observation de l'événement entre J0 et la fin du suivi
- Temps de survie = délai (en jours, mois, années) entre J0 et, soit la date de l'événement, soit la date de la censure

6.3. Interprétation des courbes de Kaplan-Meier (figure issue de Boswood, 2016)



- Axe des abscisses = délai depuis J0
- Axe des ordonnées = % estimé dans l'échantillon d'individus n'ayant pas encore présenté l'événement à un délai t depuis J0
- 1 saut de marche = survenue d'au moins 1 événement à ce délai-là depuis J0
- 1 petit tiret vertical sur une courbe = survenue d'au moins une censure à ce délai-là depuis J0 (tirets non nécessairement représentés)
- 766 jours = temps de survie médian dans le groupe placebo = en estimation, 50% des individus sous placebo n'ont pas présenté l'événement 766 jours après J0 = 50% des individus sous placebo ont présenté l'événement 766 jours après J0
- « $P = 0.0038$ » = degré de signification du test du log-rank testant si les courbes de survie sont ou non significativement différentes (dans l'exemple, elles le sont)
- Plus la pente d'une courbe de survie est marquée, plus l'événement survient rapidement après J0

7. Biais d'association dans l'estimation d'un $OR^*_{E \leftrightarrow M}$

7.1. Définition

Biais d'association = écart systématique entre l' $OR^*_{E \leftrightarrow M}$ estimé et l' $OR^*_{E \leftrightarrow M}$ dans la population cible

7.2. Biais de sélection

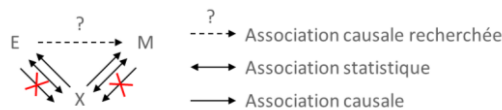
- Origine : protocole de sélection des individus dans l'étude de telle sorte que l' $OR^*_{E \leftrightarrow M}$ estimé dans l'échantillon est *a priori* différent de l' $OR^*_{E \leftrightarrow M}$ dans la population cible
- Difficile à déceler

7.3. Biais de classement

- Origine : erreurs de classement (même aléatoires) sur E ou sur M
- Biais de classement non différentiel
 - Présent si les erreurs sur E ne dépendent pas de M et si les erreurs sur M ne dépendent pas de E
 - Diminue la force de l'association → peut rendre non significative une association qui l'aurait été sans les erreurs de classement
- Biais de classement différentiel
 - Présent si les erreurs sur E dépendent de M ou si les erreurs sur M dépendent de E
 - Augmente la force de l'association → peut rendre significative une association qui ne l'aurait pas été sans les erreurs de classement → biais de classement dangereux
 - Suppression du biais de classement différentiel grâce à la mise aveugle

7.4. Biais de confusion et facteur de confusion

- Définition : écart systématique entre l' $OR^*_{E \leftrightarrow M}$ estimé et l' $OR^*_{E \leftrightarrow M}$ *causal* dans la population cible
- Origine : existence d'un facteur de confusion non pris en compte dans les analyses statistiques
- Facteur de confusion : exposition X qui répond aux deux critères ci-dessous
 - Critère n°1 : X cliniquement associée à E sans en être une conséquence ($\Leftrightarrow$ les individus E+ ne sont pas cliniquement comparables aux individus E-)
 - Critère n°2 : X cliniquement associée à M sans en être une conséquence



- Moyens de limiter les biais de confusion dus à des facteurs de confusion
 - Attribution aléatoire (randomisation) de E
 - Appariement (exemple, étude cas-témoins appariée)
 - Ajustement sur les facteurs de confusion à l'aide de modèles multivariés

8. Suggérer que E semble un facteur de risque / protecteur de M dans la population

E doit respecter la règle des 3A :

- Association significative entre E et M ($p < 0,05$)
- Valeur de E, telle que collectée dans l'étude, antérieure à la survenue de M (critère d'antériorité)
- Absence de biais de sélection, de classement différentiel, et de confusion

9. Modèles de régression multivariés

- Le choix d'un modèle de régression (multivarié) dépend du type de M
 - Si M est quantitative → régression linéaire
 - Si M est binaire sans délai de survenue de M → régression logistique
 - Si M est binaire avec délai de survenue de M → modèle de Cox
- Permettent d'ajuster l'association entre E et M sur ≥ 1 facteur de confusion en même temps → 🚫 des biais de confusion → inférence causale envisageable

10. Les principales études cliniques

10.1. Essai clinique (*controlled trial*)

- Recrutement d'individus tous indemnes de M
- Attribution (aléatoire, ou non aléatoire) de E au moment de recruter les individus
- Suivi des individus dans le temps → recueil de la date de survenue de M
- Critère d'antériorité garanti
- En général, analyses de survie à utiliser (courbes de Kaplan-Meier, modèle de Cox)

10.2. Etude observationnelle de cohorte (*cohort study*)

- Recrutement d'individus tous indemnes de M, déjà E+ ou E- à l'inclusion
- Suivi des individus dans le temps → recueil de la date de survenue de M
- Critère d'antériorité garanti
- En général, analyses de survie à utiliser (courbes de Kaplan-Meier, modèle de Cox)

10.3. Etude observationnelle cas-témoins (*case-control study*)

- Recrutement des « cas » d'un côté et des « témoins » de l'autre
- Souvent, 50% de cas et 50% de témoins (mais ce n'est pas une obligation)
- Critère d'antériorité garanti difficile à garantir pour des expositions évoluant au cours du temps
- Régression logistique ou linéaire à utiliser

10.4. Etude observationnelle transversale (*cross-sectional study*)

- Recrutement des individus plus ou moins « au hasard » dans la population source
- Critère d'antériorité garanti difficile à garantir pour des expositions évoluant au cours du temps
- Régression logistique ou linéaire à utiliser

11. Etudes randomisées versus études observationnelles

- Avantages / inconvénients des études randomisées (E attribuée aléatoirement)
 - (+) Exposition E contrôlée donc homogène (quantité, durée, observance, ...)
 - (+) Biais de confusion minimales ($\Leftrightarrow$ excellente comparabilité clinique *a priori*)
 - (+) Inférence causale « facile »
 - (-) Aspects éthique et de faisabilité pouvant rendre impossible la randomisation
- Avantages / inconvénients des études observationnelles
 - (+) Représente la « vraie vie »
 - (+) Pas de problème d'éthique ou de faisabilité
 - (-) Biais de confusion persistant (oubli / impossibilité de prendre en compte des facteurs de confusion) → inférence causale difficile ou à faire avec des pincettes
- Essais cliniques *non* randomisés : inférence causale aussi difficile à faire que dans les études observationnelles