

GLOSSAIRE

A

Ajustement de Bonferroni : méthode utilisée lors de comparaisons multiples pour éviter l'inflation du risque d'erreur α . Elle consiste à diviser le niveau de signification global souhaité par le nombre total de tests. Si par exemple il y a deux comparaisons à effectuer et que le seuil de signification global doit être de 0,05 chaque comparaison est effectuée au seuil 0,025 (0,05/2). Cette méthode, classique, est simple à utiliser. Elle n'est pour autant pas à conseiller si le nombre total de comparaisons est trop important (> 5) : la correction est telle que le remède peut devenir pire que le mal. Il vaut mieux s'orienter alors vers d'autres méthodes.

Analyse multivariée/analyse ajustée : analyse où sont évalués simultanément les effets de plusieurs variables explicatives sur une variable réponse, en général par l'utilisation d'un modèle de régression, dont le choix dépend de la nature de la variable réponse.

Analyse par landmark : méthode d'analyse utilisée pour prendre en compte les variables d'exposition dépendantes du temps.

Analyse post-hoc : analyse des données (ou d'une partie des données) d'une étude, non prévue au protocole de l'étude et réalisée *a posteriori*.

Analyse RPSFT (*rank-preserving structural failure time*) : méthode d'analyse de survie permettant de minimiser le biais lié à la dilution de l'effet d'un traitement qui survient lorsque des patients changent de bras de traitement en cours d'essai.

Analyses comparatives : ces analyses concernent le plus souvent des essais dits de « supériorité » visant à démontrer que le traitement (ou la stratégie de prise en charge) expérimental(e) est plus efficace, que le traitement (stratégie) de référence. Il existe aussi des essais de non-infériorité et d'équivalence (développement de biosimilaires et de génériques, par exemple). Dans tous les cas, la comparabilité des groupes analysés doit être assurée.

Analyses descriptives : descriptions de phénomènes liés à l'état de santé d'une population ou d'un groupe de patients tels que l'incidence, la prévalence, la mortalité, la morbidité, la létalité, ou même des variations temporelles et/ou géographiques.

Ces analyses sont informatives, mais elles ne permettent pas d'inférer des mécanismes de causalité.

Appariement (*matching*) : procédure utilisée dans les études comparatives visant à appairer les sujets entre les groupes sur des variables pré-identifiées (âge par exemple) dans le but de rendre les groupes comparables.

B

Biais d'attribution : ou biais d'indication correspondant à la non-comparabilité des facteurs pronostiques entre les groupes d'intervention qui entraîne une interprétation faussée de l'effet de l'intervention sur la survenue d'un évènement.

Biais d'attrition : biais dû aux sorties d'essai ou interruptions de traitement quand elles ne surviennent pas aléatoirement mais en fonction du bras de traitement.

Biais de sélection : différence de composition des groupes de traitements entraînant la non-comparabilité des groupes sur les facteurs de confusion. Le tirage au sort contrôle le biais de sélection entre les bras de traitement et l'analyse en intention de traiter maintient la comparabilité des groupes analysés.

Boîte de distribution (anglais : *boxplot*), dite aussi boîte à moustaches : représentation graphique d'un jeu de données sous forme de boîte. Verticale, elle se décompose en plusieurs éléments. Ligne centrale : médiane ; bord inférieur : 25^e quartile, bord supérieur : 75^e quartile. Les lignes verticales basses et hautes indiquent les centiles (5^e et 95^e) ou déciles (1^{er} et 9^e) selon option. Les données aberrantes (*outliers*) sont également souvent représentées par des points additionnels. Cette boîte peut être aussi présentée horizontalement.

Bonferroni (ajustement de) : méthode utilisée lors de comparaisons multiples pour éviter l'inflation du risque d'erreur α . Elle consiste à diviser le niveau de signification global souhaité par le nombre total de tests. Si par exemple il y a deux comparaisons à effectuer et que le seuil de signification global doit être de 0,05 chaque comparaison est effectuée au seuil 0,025 (0,05/2). Cette méthode, classique, est simple à utiliser. Elle n'est pour autant pas à conseiller si le nombre total de comparaisons est trop important (> 5) : la correction est telle que le remède peut devenir pire que le mal. Il vaut mieux s'orienter alors vers d'autres méthodes.

C

Calcul d'effectif : calcul fait au moment de la planification d'un essai afin de déterminer, en faisant certaines hypothèses sur l'effet attendu du traitement, le nombre nécessaire de patients à inclure pour obtenir la puissance voulue (c'est-à-dire la probabilité de mettre en évidence l'effet du traitement si les hypothèses sont exactes).

Calibration d'un modèle : capacité du modèle à prédire le niveau de risque d'un évènement proche du niveau de risque réel. La calibration est une des composantes de la performance de prédiction d'un modèle.

Cause-specific hazard : fonction de risque instantanée d'un évènement particulier dans le cadre des évènements en compétition.

Censure : terme utilisé dans le cadre de données de survie (cf. *survival data*). La censure est un évènement qui interrompt le suivi d'un patient, autre que l'évènement d'intérêt qui constitue le critère de jugement. Si le critère de jugement est la survie globale, l'évènement d'intérêt est le décès, et la censure peut correspondre soit à un patient perdu de vue ou bien à l'arrêt du suivi car fin de l'étude (censure administrative).

Cochran's Q test : test permettant d'évaluer l'hétérogénéité de certaines mesures dans une méta-analyse.

Cohorte : très généralement une cohorte est un ensemble de sujets. Dans une étude dite de cohorte,

les sujets sont répartis en groupes en fonction de leur exposition et l'évènement n'est pas survenu au moment où cette répartition est faite. La comparaison du taux de survenue de l'évènement entre les différentes sous-cohortes, définies par l'exposition permettra de mesurer l'association entre exposition et évènement.

Cohorte d'expansion : cohorte additionnelle de patients dans un essai de phase I exposés à la dose de traitement déterminée par un schéma d'escalade de dose. L'objectif de la cohorte d'expansion peut être d'affiner l'estimation de la toxicité, de confirmer l'identification de la DMT, d'explorer les propriétés de pharmacocinétique et de pharmacodynamique, ou de sélectionner des molécules expérimentales sur un critère d'activité/efficacité.

Complete case analysis : méthode de gestion des données manquantes où l'analyse est restreinte aux données non manquantes.

Courbe ROC : courbe permettant de représenter les capacités de discrimination d'un marqueur diagnostique ou pronostique par rapport à un critère de jugement binaire. La courbe représente, pour chaque valeur de seuil du marqueur, la sensibilité (en ordonnée) en fonction de 1-la spécificité (en abscisse).

Cox (modèle) : modèle de régression servant à mesurer l'impact d'une ou de plusieurs variables explicatives sur la distribution de survie, de façon similaire à la régression linéaire ou logistique.

Critère binaire : un critère de jugement binaire, ou dichotomique, ne peut prendre que deux modalités. Le plus souvent, il s'agit de la survenue d'un évènement durant une période de temps fixée à l'avance.

Critère composite : critère de jugement comportant plusieurs évènements, pouvant relever de l'efficacité et/ou de la tolérance de l'intervention étudiée (par exemple, décès de cause cardiovasculaire ou hospitalisation pour insuffisance cardiaque).

Critère de jugement (outcome) : critère qui permet d'évaluer l'effet du traitement dans une étude expérimentale. Le critère de jugement principal doit être

cohérent avec l'objectif de l'étude, et permet le calcul du nombre de sujets nécessaires.

Critère de substitution (*surrogate endpoint*) : critère de jugement intermédiaire pouvant être utilisé pour évaluer l'effet d'une intervention sur le critère de jugement d'intérêt sous réserve que l'effet de la maladie et l'effet de l'intervention sur la survenue du critère d'intérêt soient entièrement capturés par le critère de jugement intermédiaire. La survie sans progression est validée comme critère de substitution de la survie globale dans certaines localisations tumorales et pour certains traitements.

CRM : acronyme pour « *Continual Reassessment Method* », design adaptatif de recherche de dose en phase I introduit par O'Quigley en 1990. Sur la base d'hypothèses sur les probabilités dites a priori d'observer une toxicité dose limitante pour chaque palier à l'étude, ce schéma dit « bayésien » met à jour le calcul des probabilités a posteriori de DLT après chaque cohorte et indique ainsi la prochaine dose à explorer.

D

Degré de signification/p-value (p) : probabilité d'observer sous l'hypothèse nulle (en général hypothèse d'absence de différence) une différence au moins aussi grande que celle observée dans l'étude. Lorsqu'il est inférieur au risque de première espèce (fixé en général à 0,05), on conclut à une différence statistiquement significative. Le degré de signification quantifie la force de notre conviction : plus il est faible, moins on pense que la différence observée dans l'étude puisse être liée au hasard.

Discrimination d'un modèle : capacité du modèle à différencier les patients à haut risque des patients à bas risque de développer un événement. La discrimination est une des composantes de la performance de prédiction d'un modèle.

E

Échantillon : sous-ensemble d'une population sélectionnée pour les besoins d'une étude.

Erreur de type 1 : probabilité de rejeter à tort l'hypothèse nulle alors qu'elle est vraie.

Essais séquentiels : essais cliniques dans lesquels une ou plusieurs analyses intermédiaires ont été planifiées pour permettre de stopper les inclusions précocement ou non en fonction des premiers résultats.

Estimation de Mantel-Byar : méthode d'analyse utilisée pour prendre en compte les variables d'exposition dépendantes du temps.

Étude cas-témoin (*case control study*) : étude rétrospective observationnelle dans le but d'évaluer l'association entre des facteurs de risques (exposition) et une maladie.

Étude pangénomique : étude évaluant l'association entre des polymorphismes génétiques et une maladie.

Étude prospective : étude dans laquelle l'exposition est mesurée avant la survenue de l'évènement étudié.

Étude randomisée : étude dans laquelle l'attribution du traitement a été tirée au sort. Le tirage au sort garantit l'équilibre, en moyenne, de tous les facteurs pronostiques connus et inconnus entre les groupes de traitement. Dans une étude randomisée idéale on peut estimer que les patients traités par A et B sont échangeables ce qui permet l'affirmation du lien de causalité entre différence de devenir des deux populations et traitement.

Étude rétrospective : étude dans laquelle l'exposition est mesurée après la survenue de l'évènement étudié.

Event free survival (EFS)/ survie sans évènement : analyse de survie où l'on s'intéresse au délai de survenue d'un évènement composite regroupant le décès et la survenue d'un évènement (rechute, complication, en fonction de la définition choisie). En pratique on s'intéressera au délai du plus précoce de ces différents évènements.

F

Facteur de confusion : Facteur ou variable lié à la fois à la variable à expliquer (critère de jugement par exemple) et la variable explicative (traitement). Peut être responsable d'un biais de confusion pour l'analyse de l'association entre la variable explicative et la variable à expliquer.

Facteur de risque : variable associée à la survenue d'une maladie.

Facteur pronostique : variable associée à l'évolution (décès, survenue d'une complication) d'une maladie.

Fine and Gray (modèle) : Modèle de régression non pas basé sur la fonction de risque instantané cause-spécifique (modèle de Cox) mais sur la fonction de risque instantané associée à la fonction d'incidence cumulée (appelée fonction de risque de sous-répartition). Dans sa forme classique, le modèle de *Fine and Gray* considère une formulation à risques proportionnels.

Flowchart / diagramme de flux : organigramme décrivant le recrutement, la randomisation, et le suivi de l'ensemble des patients lors de l'étude.

Forest plot : terme désignant à l'origine une représentation graphique des résultats de plusieurs études dans une méta-analyse ou une revue systématique. En pratique, il est utilisé en statistique pour illustrer les différents résultats d'une même analyse (régression logistique, modèle de Cox notamment) répétée au sein de sous-groupes différents. Les points centraux représentent les valeurs estimées des *odds* ou *hazard ratios* et les lignes horizontales les intervalles de confiance associés. On représente avec une ligne verticale la valeur-seuil (0 pour les OR, 1 pour les HR) ; ainsi, un intervalle de confiance ne contenant pas ce seuil révèle un effet significatif. L'emploi du terme est donc abusif et on doit plutôt le considérer comme un « *hazard ratios plot by subgroups* ».

H

Hazard Ratio : rapport des risques instantanés, mesure d'effet utilisée dans les modèles de régression pour données de survie.

I

Imputation multiple : méthode de gestion des données manquantes.

Incidence : nombre de nouveaux cas observés dans une population donnée, divisé par la taille de cette population et la durée de la période d'observation, à ne pas confondre avec la prévalence.

Index I² : quantifie le degré d'hétérogénéité dans une méta-analyse. Un I² à 0 signifie une variabilité conforme à celle liée à l'échantillonnage.

Index C (Harrell) : mesure de concordance permettant d'évaluer la discrimination d'un modèle ou d'un marqueur.

Incidence : nombre de nouveaux cas observés dans une population donnée, divisé par la taille de cette population et la durée de la période d'observation, à ne pas confondre avec la prévalence.

Intention de traiter : type d'analyse considérant les patients dans le groupe de traitement dans lequel ils ont été randomisés initialement, quel que soit le traitement effectivement reçu, leur devenir dans l'essai, leur observance au traitement alloué ou les éventuels retraits de l'étude ou déviations au protocole. Souvent désignée avec l'acronyme « ITT » pour « *Intent To Treat* ».

K

Kaplan-Meier (méthode) : estimation non paramétrique de la fonction de survie $S(t)$ = probabilité d'être encore en vie au temps t qui permet de tenir compte de la censure, c'est-à-dire de l'interruption

du suivi de certains patients avant la survenue de leur décès.

La représentation de la survie au cours du temps est appelée courbe de Kaplan-Meier.

L

Landmark (analyse par) : méthode d'analyse utilisée pour prendre en compte les covariables dépendantes du temps (exemple la rechute). Cette méthode permet d'éviter l'analyse biaisée qui consiste à différencier deux groupes, les répondeurs et les non-répondeurs et de comparer leur survie.

M

Marqueur prédictif : marqueur associé à la réponse au traitement. Le traitement et le marqueur prédictif interagissent sur le devenir du patient. L'interaction peut être qualitative (réponse observée ou non selon l'expression du marqueur) ou quantitative (intensité de la réponse dépendant de l'expression du marqueur).

Marqueur pronostique : marqueur associé au devenir du patient, à la survenue du critère de jugement d'intérêt.

Médiane de suivi : délai au-delà duquel moins de 50 % des patients sont encore suivis.

Médiane de survie globale (median OS) : délai au-delà duquel 50 % des patients sont décédés.

Médiane de survie sans progression (median PFS) : délai au-delà duquel 50 % des patients sont décédés ou ont vu leur maladie progresser.

Mesure poolée : se dit de plusieurs mesures combinées entre elles, par exemple en utilisant la moyenne de plusieurs mesures.

Méta-analyse : démarche statistique combinant les résultats de plusieurs études indépendantes sur une question donnée.

Méthode de réévaluation séquentielle (MRS) : schéma d'étude de phase I.

Modèle à effets aléatoires : un modèle à effet aléatoire est un modèle dont certains paramètres sont caractérisés par une distribution (par exemple une distribution normale) et pas par une valeur fixe. Cela permet par exemple de tenir compte d'un effet centre.

Modèle de Cox : modèle de régression utilisé pour l'analyse de délais d'évènement.

Modified Toxicity Probability Interval (mTPI) : méthode d'escalade de dose en phase I introduite par Ji et Wang. À partir d'une probabilité cible de toxicité limitant la dose (DLT), la méthode établit des règles d'exploration des doses en fonction du coût des décisions suivantes : désescalader d'un palier de dose, rester dans le palier de dose, escalader d'un palier de dose.

N

Non-infériorité : essai visant à démontrer qu'un traitement expérimental a une efficacité suffisante comparativement à une thérapeutique standard, c'est-à-dire que la perte d'efficacité est inférieure à un seuil défini comme acceptable.

O

Odds ratio : « rapport des chances » ou « rapport des cotes » en français, qui mesure le degré de dépendance entre des variables qualitatives et permet de comparer les probabilités d'observer un événement entre deux groupes dans le cadre d'une régression logistique.

P

Per protocol : type d'analyse considérant uniquement les patients ayant effectivement reçu le traitement



prévu par la randomisation et traités en pleine conformité avec le protocole. Souvent désignée avec l'acronyme « PP ».

Placebo : préparation dépourvue de tout principe actif.

Plans de Simon : de nombreux plans adaptés aux essais thérapeutiques de phase 2, à un seul groupe de traitement ont été proposés (Gehan, Fleming, Ensign, et bien d'autres...). Parmi ceux-ci figurent ceux de R. Simon. Ses plans se déclinent en deux étapes, avec pour principe de permettre un arrêt précoce de l'essai dès la première étape si des résultats observés sont déjà en faveur de l'efficacité ou de l'inefficacité du produit étudié. Il existe deux variantes :

- le plan dit « Optimal » qui minimise le nombre moyen de sujets exposés ;
- le plan dit « Minimax » qui minimise le nombre maximum de sujets exposés.

Prévalence : nombre de cas de maladies présents à un moment donné (que le diagnostic soit récent ou ancien) divisé par la taille de la population de l'étude. Correspond donc à la proportion de sujets malades à un instant donné dans la population.

Progression free survival (PFS)/survie sans progression : analyse de survie où l'on s'intéresse au délai de survenue d'un évènement composite regroupant le décès et la progression de la maladie (*i.e.* au délai du plus précoce de ces deux évènements).

Progression free survival 2 (PFS2)/survie sans progression 2 : critère de jugement où l'on s'intéresse au délai de survenue d'un évènement composite regroupant le décès et la progression de la maladie après une deuxième ligne de traitement. La progression sous la 1^{re} ligne de traitement n'est pas donc pas considérée comme un évènement avec ce critère.

Puissance : dans le cadre d'un test statistique, probabilité de mettre en évidence une différence entre deux populations, à partir d'une étude sur échantillons de patients, sachant qu'il existe réellement une différence.

Puissance conditionnelle : calcul de puissance effectué une fois l'essai débuté et tenant compte des premiers résultats.

R

Randomisation : affectation aléatoire des participants dans les groupes de traitements.

Recul : délai entre la date de début du suivi et la date de l'analyse (ou date d'arrêt du recueil de données de l'étude). Le recul est individuel, spécifique à chaque patient.

Régression logistique : modèle de régression linéaire généralisé mesurant l'effet d'un groupe (vecteur) de variables aléatoires sur une variable aléatoire binomiale. Son extension à des variables qualitatives à plus de deux modalités est la régression multinomiale.

Risque alpha/risque de première espèce : probabilité de rejeter à tort l'hypothèse nulle (en général hypothèse d'absence de différence) alors qu'elle est vraie. Ce seuil est de manière consensuelle fixé dans la littérature biomédicale à 5 %. Si le degré de signification (*p-value*) du test est inférieur au risque alpha, on rejette l'hypothèse nulle et on conclut à une différence significative.

Risque bêta/risque de deuxième espèce : probabilité de ne pas conclure à une différence alors que cette différence existe. Ce risque n'a de sens qu'en fonction de la différence que l'on souhaite mettre en évidence. Dans une étude on fixe la puissance que l'on veut obtenir pour calculer le nombre de sujets nécessaires pour mettre en évidence la différence minimale cliniquement pertinente. On choisit souvent une puissance de 90 %, donc un risque bêta de 10 % (puissance = 1 - bêta)

Risques compétitifs ou en compétition (*competing risks*) : situation d'analyse de survie où le patient peut être exposé simultanément à la survenue de plusieurs évènements, la survenue de l'un annulant ou modifiant la probabilité de survenue des autres.

Risques instantanés : cette fonction s'appelle aussi force de mortalité ou fonction de risque. C'est une approximation de la probabilité de survenue de l'évènement dans un intervalle de temps. Il faut noter que la probabilité d'évènement à chaque instant t est infiniment petite.

RMST (Restricted Mean Survival Time) : méthode d'analyse de survie utilisée en l'absence de proportionnalité des risques.

S

Schémas « 3+3 » : schéma d'étude de phase I consistant à administrer une certaine dose de médicament à 3 patients et à observer la fréquence des toxicités. Si aucune toxicité n'est observée on administre aux 3 patients suivants une dose plus importante, si une toxicité est observée on administre de nouveau la même dose aux 3 patients suivants, si plus d'une toxicité est observée on s'arrête là.

Schémas « A+B » : généralisation du schéma « 3+3 » à des cohortes de taille A et B.

Schémas up and down : généralisation du schéma « 3+3 » permettant de diminuer de dose en cas d'observation de plus d'une toxicité sur la cohorte de 3 patients.

Score de propension : probabilité de recevoir un traitement, issue d'un modèle de régression sur les caractéristiques initiales. Ce score permet de neutraliser le biais d'attribution dans les études observationnelles.

Score pronostique : score permettant d'estimer le pronostic d'une personne. Il est établi selon plusieurs facteurs pronostiques.

Stratification : terme employé pour indiquer la prise en compte d'un facteur de confusion. S'emploie

soit dans le cadre du design d'un essai randomisé (randomisation stratifiée) soit au moment de l'analyse statistique des données.

Survie médiane : délai au-delà duquel 50 %, des patients de l'étude sont décédés (ou ont expérimenté l'évènement d'intérêt si l'*outcome* n'est pas le décès).

Survie relative : compare la survie chez un groupe de personnes atteintes d'une pathologie à la survie prévue pour un groupe de personnes dans la population générale qui présentent les mêmes caractéristiques que les personnes atteintes de cette pathologie, telles que les antécédents, le sexe, l'âge... Ainsi, un taux de survie relative après 5 ans de 90 % signifie qu'en moyenne, les personnes atteintes de cette pathologie ont 90 % de chances de vivre 5 ans ou plus après avoir reçu leur diagnostic comparativement aux personnes du même âge et du même sexe qui font partie de la population générale.

Survival data/Time to event data / Time to failure data : données pour lesquelles la variable d'intérêt est un délai d'évènement (event/failure) : délai jusqu'au décès pour la survie, ou délai jusqu'à un évènement d'intérêt ou jusqu'au premier de plusieurs évènements d'intérêt en fonction des définitions choisies (EFS, PFS). Le fait que les délais d'évènement aient une distribution asymétrique et qu'une partie des données soit censurée (l'évènement n'est pas observé pour la totalité des patients de l'étude) impose d'utiliser des méthodes d'analyse spécifiques.

T

Test de Gray : test de comparaison des fonctions d'incidence cumulée.

Test du Logrank : test de comparaison des fonctions de survie.