

LEXIQUE
de la
recherche clinique
et de la
médecine factuelle
(Evidence-Based Medicine)

K. Slim

Service de chirurgie générale et digestive, Clermont-Ferrand



ELSEVIER
MASSON



Ce logo a pour objet d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, tout particulièrement dans le domaine universitaire, le développement massif du «photocopiage». Cette pratique qui s'est généralisée, notamment dans les établissements d'enseignement, provoque une baisse brutale des achats de livres, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

Nous rappelons donc que la reproduction et la vente sans autorisation, ainsi que le recel, sont passibles de poursuites. Les demandes d'autorisation de photocopier doivent être adressées à l'éditeur ou au Centre français d'exploitation du droit de copie : 20, rue des Grands-Augustins, 75006 Paris, Tél. : 01 44 07 47 70.

Correspondance

Karem Slim, Service de chirurgie générale et digestive,
Hôtel-Dieu Boulevard Léon Malfreyt F-63058 Clermont-Ferrand
e-mail : kslim@chu-clermontferrand.fr

Tous droits de traduction, d'adaptation et de reproduction par tous procédés, réservés pour tous pays.

Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit, des appes publiées dans le présent ouvrage, faite sans l'autorisation de l'éditeur est illicite et constitue une contrefaçon. Seules sont autorisées, d'une part les reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisaton collective, et d'autre part, les courtes citations justifiées par le caractère scientifique ou d'information de l'œuvre dans laquelle elles sont incorporées (art. L. 122-4, L.122-5 et L. 335-2 du Code de la propriété intellectuelle).

© 2007, Elsevier Masson SAS. Tous droits réservés
ISBN : 978-2-294-70400-0

Préface

Les progrès en médecine sont le fruit d'innovations, mais aussi d'évaluations. Sans innovations il n'y aurait pas de progrès. Mais il est nécessaire de s'assurer, par des évaluations scientifiques, qu'une innovation apporte réellement un progrès. En effet, l'histoire de la médecine montre que des impressions subjectives concernant certains moyens diagnostiques ou certains traitements ont été prises pour des progrès ce qui a été infirmé par la suite : des croyances ou des espoirs ne sont pas forcément des réalités.

Depuis la deuxième moitié du xx^e siècle, des méthodes d'évaluation se sont développées et ont acquis peu à peu droit de cité dans la recherche médicale. La connaissance de ces méthodes devient indispensable pour les médecins afin qu'ils puissent se faire une opinion critique personnelle sur les travaux publiés. Ainsi, peuvent-ils et doivent-ils fonder leurs prescriptions sur les publications dont les résultats et leur interprétation reposent sur les meilleurs niveaux de rigueur scientifique.

À ces techniques d'évaluation, relativement nouvelles en médecine, s'associe un langage nouveau. Le mérite de ce lexique est d'offrir à tous les membres des professions de santé, comme aux étudiants qui ne sont pas familiarisés avec les techniques méthodologiques actuelles, des clés pour comprendre facilement ce qui, sinon, pourrait leur paraître ésotérique.

Aussi, ce livre devrait-il être accessible à tous, de la même façon qu'il l'est par sa conception.

Professeur Michel Huguier
Service de chirurgie digestive
Hôpital Tenon, Paris

Abréviations

AAR :	augmentation absolue du risque
CEBM :	<i>Centre for Evidence-Based Medicine</i>
CENTRAL :	<i>Cochrane Central Register of Controlled Trials</i>
CONSORT :	<i>Consolidated Standards of Reporting Trials</i>
EBM :	<i>Evidence-Based Medicine</i>
ECR :	essai contrôlé randomisé
ET :	écart type, erreur type
GMI :	<i>Growth Modulation Index</i>
HAS :	Haute Autorité de Santé
IC :	intervalle de confiance
ISI :	<i>Institute for Scientific Information</i>
JCR :	<i>Journal Citation Reports</i>
MeSH :	<i>Medical Subject Heading</i>
MOOSE :	<i>Meta-analysis of Observational Studies in Epidemiology</i>
NICE :	<i>National Institute for Clinical Excellence</i>
NLM :	<i>National Library of Medicine</i>
NNH :	<i>Number Needed to Harm</i>
NNN :	nombre nécessaire pour nuire
NNT :	<i>Number Needed to Treat</i>
NST :	nombre de sujets à traiter
OR :	odds ratio
PICO :	<i>Patient Intervention Control Outcome</i>
QALY :	<i>Quality Adjusted Life Years</i>
QdV :	qualité de vie
QUOROM :	<i>Quality of Reporting of Meta-analysis</i>
RAR :	réduction absolue du risque
RC :	rapport des cotes
RCT :	<i>Randomized Controlled Trial</i>
RM :	revue méthodique de la littérature
ROC :	<i>Receiver Operating Characteristic</i>
RPC :	règle de prescription clinique, recommandation de pratique clinique
RR :	risque relatif
RRR :	réduction relative du risque
RV :	rapport de vraisemblance
SEM :	<i>Standard Error of the Mean</i>
SIR :	<i>Standardized Incidence Ratio</i>
SMR :	<i>Standardized Mortality Ratio</i>
STARD :	<i>Standards for Reporting of Diagnostic Accuracy</i>
STROBE :	<i>Strengthening the Reporting of Observational Studies in Epidemiology</i>
TAE :	taux attendu de l'événement
TREND :	<i>Transparent Reporting and Evaluation with Non-randomized Designs</i>
VPN :	valeur prédictive négative
VPP :	valeur prédictive positive

Introduction

Cher lecteur,

Ce lexique de la recherche clinique et de la médecine factuelle est réalisé pour répondre à une demande légitime des praticiens (jeunes et moins jeunes) et des médecins en formation qui sont confrontés de manière croissante (dans leurs lectures ou dans leurs participations aux congrès) à l'usage d'un jargon scientifique parfois « opaque ». Le but de cet ouvrage est de rendre cette terminologie accessible au plus grand nombre.

La méthodologie de la recherche clinique est souvent discutée dans les colonnes des revues scientifiques. Cette recherche dispose de ses propres codes, que les praticiens doivent connaître pour ne pas manquer le train d'un concept moderne appelé « médecine factuelle ». Selon ce nouveau paradigme, l'expérience personnelle et le sens clinique ne doivent plus être les seuls éléments pour guider le médecin (ou le chirurgien) dans ses choix thérapeutiques. Le recours aux meilleurs produits de la recherche clinique et l'interprétation éclairée de leurs résultats sont essentiels.

La médecine fondée sur les meilleures preuves scientifiques, ou médecine factuelle (ou *Evidence-Based Medicine*), que certains ont qualifié de « révolution silencieuse », n'est pas une mode passagère mais un concept définitivement établi dans la pratique quotidienne.

Il est donc indispensable pour nous, praticiens confrontés de manière constante à des situations cliniques diverses, d'assimiler la signification exacte des termes utilisés dans la section « méthodes » des articles scientifiques, de discerner les publications fondées sur les meilleures preuves, de connaître les recommandations factuelles de pratique clinique, ou plus globalement de suivre et comprendre la recherche clinique.

Il est nécessaire de savoir si les conclusions proviennent d'études cas-témoins ou d'études randomisées, si l'essai est en double ou simple aveugle, si les tests diagnostiques utilisés ont de bonnes sensibilité et spécificité, si la méta-analyse est de bonne qualité, quelle est la différence entre risque absolu et risque relatif, quelle est la pertinence clinique de l'odds ratio.

Pour que la méthodologie et la recherche ne restent pas des domaines hermétiques et réservés à un « cercle fermé », les lecteurs de ce lexique auront à leur disposition les définitions des termes les plus

fréquemment rencontrés. Chaque expression (avec sa traduction anglaise, puisque la majorité de nos lectures est en anglais) est définie de manière claire et concise. Le cas échéant, un exemple pratique est fourni. Un renvoi aux sources Internet est donné chaque fois que cela semble pouvoir aider les lecteurs à approfondir le sujet. Les sources consultées pour l'élaboration de ce glossaire sont multiples. Je peux citer : Le grand dictionnaire terminologique (www.granddictionnaire.com), Méthodologie et interprétation des essais thérapeutiques (www.spc.univ-lyon1.fr/polycop), *Glossaire des termes utilisés en Evidence-Based Medicine* (Van Driel M., éd. Minerva, revue d'EBM, Gent, Belgique, 2004), *Méthodes statistiques à l'usage des médecins et des biologistes* (Schwartz D., éd. Flammarion Médecine-Sciences, Paris, 1993), *Biostatistiques au quotidien* (Huguier M., Flahault A., éd. Elsevier, Paris, 2003), et des bases de données électroniques telles que www.google.fr ou www.scholar.google.com. À la fin de l'ouvrage, les lecteurs disposent aussi de la liste alphabétique des traductions anglaises de tous les termes expliqués dans ce lexique.

Remerciements

Plusieurs amis ont eu l'obligeance de relire ce texte et me faire part de leurs critiques ou commentaires. Je les en remercie très cordialement.

Fabrice Kwiatkowski, tu as vérifié la définition des termes statistiques, tu as fait preuve depuis des années d'une disponibilité désintéressée et sans faille, je ne t'en remercierai jamais assez.

Jacques Chipponi, vous êtes mon Maître, vous m'avez accepté à vos côtés, ma gratitude n'a d'égal que la profonde estime que j'éprouve à votre égard.

Je n'oublierai pas : Olivier Aumâtre, Jean Etienne Bazin, et Christian Dualé de Clermont-Ferrand, Michel Huguier et François Lacaine de Paris, et Chadly Dziri de Tunis qui ont attentivement relu ce lexique et ont grandement participé à l'améliorer.

Un petit clin d'œil à Jean-Pierre Cardin qui m'a donné l'idée de rédiger ce livre en posant les premières questions pertinentes.

Biographie



Karem Slim est chirurgien des hôpitaux et professeur associé en chirurgie digestive. Déjà rédacteur en chef aux *Annales de chirurgie*, il est depuis 2007 rédacteur en chef pour les articles originaux du *Journal de chirurgie*. Il a publié (ou a été coauteur de) plus de 200 articles scientifiques (www.scopus.com), dont près de 50 dédiés à la médecine fondée sur les preuves scientifiques (médecine factuelle ou *Evidence-Based Medicine*). Il a aussi publié un ouvrage intitulé *Pratiques validées en chirurgie* (Paris : Elsevier, 2004).

a

ACTUARIELLE (SURVIE)

Actuarial (survival)

Une étude de survie sert à calculer la variation, au cours du temps, de la probabilité de survie des patients à partir de la date d'entrée dans une étude. Le calcul de la survie actuarielle consiste à compter (à certaines dates) le nombre de personnes encore en vie, décédées (à cause de la pathologie étudiée) ou perdues de vue.

Les patients ne rentrant pas tous simultanément dans l'étude (mais de manière progressive au fil des inclusions), on n'aura pas, au moment de l'interprétation des données, le même recul pour tous les patients. De la même manière que pour la courbe de Kaplan-Meier, les méthodes de calcul actuariel utilisent l'information fournie par les patients à faible recul.

Pour la méthode actuarielle, on calcule à chaque intervalle de temps défini *a priori* (contrairement à l'estimation de Kaplan-Meier), la probabilité de mourir à chaque période en divisant le nombre de patients morts pendant cette période par le nombre de patients présents au début de la période.

Par exemple, sur 200 patients traités, 140 sont en vie à 1 an (soit 70 %), et 100 le sont à 2 ans (100/140, soit 71,4 %). On obtient ainsi une courbe qui, à la différence de la courbe de survie de Kaplan-Meier, fixe des intervalles réguliers (figure 1). Si l'effectif de l'étude de survie est faible, l'interprétation de la courbe doit être prudente du fait du risque d'erreurs statistiques.

➔ **ESTIMATION DE KAPLAN-MEIER**

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

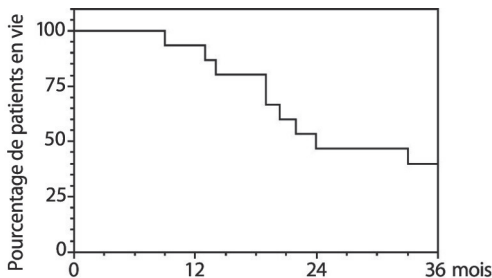


Figure 1. Représentation schématique d'une courbe de survie actuarielle.

AJUSTEMENT

➔ STANDARDISATION

ALGORITHME

Algorithm

Description des différentes étapes de la prise en charge de patients. À chaque étape, on décide, sur la base des données scientifiques les plus pertinentes et du contexte du patient, de la meilleure approche diagnostique ou thérapeutique à mettre en œuvre.

ALPHA (RISQUE)

Alpha risk

Le risque α , ou risque de première espèce (risque de type 1), est celui de conclure à tort qu'un examen, un traitement, ou un facteur pronostique est différent d'un autre, alors que cela est simplement dû au hasard. Les tests statistiques usuels limitent généralement ce risque à 5 %, seuil toléré par les méthodologistes. Ce seuil est rectifié à la baisse (divisé par le nombre de comparaisons prévues) en cas de comparaisons multiples (correction de Bonferroni).

➔ STATISTIQUEMENT SIGNIFICATIF

AMPLEUR DE L'EFFET (STANDARDISÉ)

Effect size (standardised)

L'ampleur de l'effet est le principal résultat des méta-analyses (analyse regroupant plusieurs études). Dans une méta-analyse, les données des études ne sont pas simplement additionnées (*pooled*) mais font l'objet d'un calcul spécifique en introduisant un facteur de pon-

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

dération. Pour les variables binaires (ou dichotomiques, par exemple oui/non), le résultat de la comparaison entre les deux groupes (groupe contrôle et groupe traité) est exprimé en risque relatif (RR), *odds ratio* (OR) ou différence de risque. Si les résultats concernent une même unité, on parle d'ampleur de l'effet. Si les unités sont différentes, une standardisation est nécessaire, on parle alors d'effet standardisé. En cas de variables continues, l'ampleur de l'effet est exprimée en différence de moyennes.

L'ampleur de l'effet est le reflet de la différence entre les résultats des groupes comparés. Par exemple, dans une méta-analyse sur la survie, un RR = 2 en faveur du groupe traité signifie que l'ampleur de l'effet est de 2 ; en d'autres termes, le groupe traité a deux fois plus de chances de survie que le groupe contrôle.

La figure 2 (sur les risques de la préparation mécanique du côlon) montre une ampleur de l'effet de 1,7 (losange en bas de la figure, en faveur du groupe contrôle), soit 70 % de risque supplémentaire de fistule anastomotique dans le groupe ayant une préparation colique.

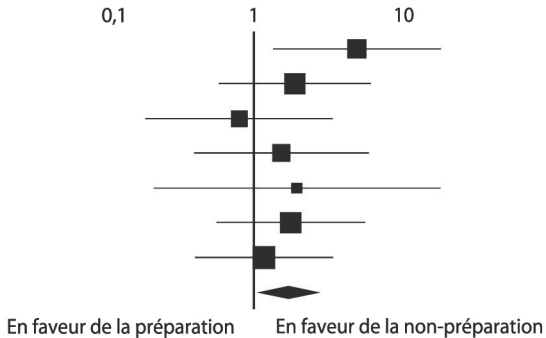


Figure 2. Exemple de méta-analyse permettant de voir l'ampleur de l'effet.

ANALYSE DE SENSIBILITÉ

Sensitivity analysis

Cette analyse consiste à voir dans quelle mesure les résultats de l'étude sont influencés par le changement de méthode, de variables, ou de critères de jugement. Plusieurs types de scénarios sont analysés.

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

Exemple

Dans une méta-analyse, l'analyse de sensibilité peut comporter un deuxième calcul de l'ampleur de l'effet après exclusion des essais de mauvaise qualité, ou bien le calcul de l'ampleur de l'effet dans certains sous-groupes. La figure 2 (voir *supra*) montre une analyse de sous-groupes avec une ampleur de l'effet différente selon le sous-groupe considéré : effet non significatif dans le premier sous-groupe incluant 3 essais, et significatif dans le deuxième sous-groupe incluant 4 essais.

ANOVA

➔ VARIANCE (ANALYSE DE)

APPARIEMENT

Matching

C'est une méthode destinée à rendre comparable la distribution de deux ou plusieurs groupes selon des caractéristiques comme l'âge, le sexe, les antécédents, ou le stade d'une tumeur. L'avantage de cette méthode est d'éviter que ces caractéristiques ne biaisent des résultats et ne rendent leur interprétation erronée, mais l'inconvénient est l'impossibilité d'analyser ultérieurement leur effet. Le groupe contrôle est généralement historique, mais un nouveau groupe contrôle peut être constitué dans le cadre de l'étude pour plusieurs raisons : soit que le groupe historique ne respecte pas les critères d'appariement, soit qu'il n'est pas suffisamment étudié (paramètres manquants), soit parce que le contexte de sa constitution (antériorité, géographie, environnement socioculturel) est trop éloigné de l'étude actuelle.

APPRENTISSAGE (COURBE D')

➔ COURBE D'APPRENTISSAGE

ASKMEDLINE

➔ PICO

AVEUGLE

Blind

Une étude est menée en aveugle lorsque les participants et/ou les personnes interprétant les résultats ignorent les traitements alloués.

On parle aussi d'« insu ». Le but de cette méthode est de garantir l'objectivité dans l'interprétation des résultats.

Il peut s'agir d'un « simple aveugle », ou « simple insu », lorsque le médecin est au courant du type de traitement mais pas le patient. Dans ce cas, il est souhaitable que la personne devant collecter ou interpréter les données ne soit pas au courant du traitement.

On parle d'un « double aveugle », ou « double insu », lorsque ni le patient ni le médecin délivrant le traitement ou réalisant l'intervention chirurgicale ne sont au courant du type de traitement alloué.

Les essais comparant deux techniques chirurgicales ne peuvent pas être menés en double aveugle (le chirurgien sait obligatoirement l'intervention qu'il réalise).

Un exemple d'étude randomisée chirurgicale menée en simple aveugle est l'essai de Majeed *et al.* ayant comparé la cholécystectomie par laparoscopie à la cholécystectomie par minilaparotomie (*Lancet* 1996, 347 : 989-94).

Un exemple d'impossibilité de mener une étude en aveugle est un essai comparant la conservation sphinctérienne à l'amputation rectale dans le traitement des cancers du bas rectum (la présence d'une colostomie « visible » empêche toute évaluation en aveugle).

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

b

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

BAYES (ANALYSE OU PROBABILITÉ BAYESIENNE)

Bayesian analysis

Le théorème de Bayes renseigne sur la probabilité conditionnelle, c'est-à-dire la probabilité que se produise un événement B sachant qu'un événement A a déjà eu lieu. Par exemple, la probabilité qu'une personne soit atteinte d'une maladie en présence d'un signe clinique (ou d'un test positif). Le calcul tient compte de la prévalence de la maladie et de la sensibilité du signe clinique ou du test. Quand elle est appliquée à la décision clinique, la formule de Bayes est généralement exprimée en *odds*. La relation de Bayes est appliquée par extension aux valeurs prédictives positives et négatives d'un signe clinique ou d'un test diagnostique.

Exemple : analyse bayésienne d'un test hémocult

En partant d'une prévalence des cancers colorectaux de 0,3 %, d'une sensibilité du test de 50 %, et d'un taux de faux positifs de 3 %, on peut calculer la probabilité qu'une personne ayant un test positif ait réellement un cancer colorectal. Le résultat de la formule de Bayes est ici de 0,05.

BÊTA (RISQUE)

Beta error, type 2 error

Le risque bêta (ou erreur β , ou erreur de type 2, ou de deuxième espèce) est celui de conclure à tort qu'un test diagnostique ou un traitement n'est pas différent d'un autre (hypothèse nulle), alors qu'il l'est en réalité. Ce risque est d'autant plus élevé que le nombre de patients comparés est faible. Le corollaire en est qu'une étude « négative » qui n'a pas tenu compte de ce risque (en calculant le nombre de patients nécessaire) est critiquable et souvent rejetée par les comités de rédaction des revues médicales.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

La puissance d'un test (ou d'une étude) est le complément du risque de type 2 ($1 - \beta$). Par exemple, si $\beta = 10\%$, la puissance est de 90% . Le seuil du risque β est souvent fixé de manière arbitraire à 10% voire à 20% (si l'enjeu médical le permet). Toutefois, si une puissance inférieure à 90% peut paraître acceptable lors d'un essai de différence (on s'attache à prédire à partir de l'échantillon une différence significative entre les groupes), elle doit être plus élevée (de 95 à 99%) lors d'un essai d'équivalence (ce traitement nouveau est mieux toléré, mais est-il aussi efficace ?).

Une fois les risques α et β fixés, on peut calculer l'effectif de l'échantillon à inclure dans une étude.

BIAIS

Biais

Les biais sont les erreurs qui affectent toute observation (on parle alors d'erreur systématique, ou *systematic error*), qui font que les résultats obtenus sont différents de la réalité. Les biais peuvent se situer à chacun des stades de la recherche : conception et élaboration de l'étude, collecte des résultats, analyse des résultats, interprétation de l'analyse, publication de l'étude.

Les types de biais rencontrés en recherche sont multiples :

- biais de sélection : la population de l'étude ne correspond pas au monde réel, les personnes incluses ne sont pas représentatives de la population générale et/ou de la maladie étudiée, il y a eu auto-sélection ;
- biais de puissance : l'échantillon sélectionné est trop faible pour autoriser une prédiction sur la population concernée ;
- biais d'indication : les choix thérapeutiques ne correspondent pas aux bonnes pratiques cliniques ;
- biais d'allocation : les traitements comparés ne sont pas répartis de manière aléatoire, l'application du protocole diffère selon les groupes ;
- biais de traitement : des sujets que l'on croit traités ne le sont en fait pas et sont des contrôles, ou un autre facteur a modifié sensiblement les paramètres mesurés ;
- biais de suivi ou de réalisation : pertes de documents, perdus de vue ;
- biais de mesure : techniques de mesure incorrectes ou non référencées, définition équivoque des critères de jugement, erreurs d'observation, critères non évalués ou oubliés ;
- biais d'attrition : retrait de certains patients de l'analyse ;

- biais d'observation ou de détection : évaluation subjective et critique des critères de jugement ;
- biais d'interprétation : erreur dans le mode d'analyse des résultats, méthode statistique inadaptée ou erronée ;
- biais de publication ou d'autocomplaisance : les données sont davantage diffusées lorsqu'elles « arrangent » les auteurs de l'étude : forme non négligeable de biais où les études dites « négatives » sont souvent refusées par les revues ;
- biais de spectre : en médecine, un test diagnostique peut offrir une sensibilité qui varie en fonction de la gravité de la maladie.

BILATÉRAL (TEST)

Two tailed ou two sided analysis

Le test bilatéral est utilisé lorsqu'on évoque un risque de troisième espèce en plus des risques α et β . Le risque de troisième espèce (qui reste relativement rare), appelé aussi risque γ , consiste à conclure à tort à la supériorité d'un test diagnostique ou d'un traitement alors qu'en réalité il est inférieur (ou inversement). Les trois types de risques d'erreur sont résumés dans le tableau I.

Quand l'hypothèse nulle est rejetée, on conclut que les tests ou les traitements sont différents mais c'est le résultat qui montrera lequel est supérieur. Le test bilatéral implique l'inclusion d'un plus grand nombre de patients.

Tableau I. Concordance entre les résultats statistiques et la réalité montrant les différents types d'erreurs statistiques.

Résultats statistiques	La réalité (inconnue)		
	A > B	A = B	A < B
A > B	Correct	Erreur de type 1 (α)	Erreur de type 3 (γ)
A = B	Erreur de type 2 (β)	Correct	Erreur de type 2 (β)
A < B	Erreur de type 3 (γ)	Erreur de type 1 (α)	Correct

(D'après Huguier M. et Flahault A. *Biostatistiques au quotidien*. Paris : Elsevier, 2003.)

BLAND ET ALTMAN (TEST DE)

Bland-Altman test

Dans l'évaluation d'une méthode diagnostique, un nouveau test est idéalement comparé à un *gold standard* (ou étalon). Mais dans certains cas, on peut être confronté à l'absence de *gold standard* ;

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

alors, plutôt que d'utiliser une régression linéaire, on fait appel au test de Bland et Altman (Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986, 1 : 307-10). Il s'agit de représenter, pour chaque cas, la différence moyenne entre les valeurs estimées issues des deux méthodes comparées en fonction de la moyenne. La moyenne des différences représente le biais moyen entre les deux méthodes. Si les différences suivent une loi normale (ou gaussienne), 95 % des différences seront comprises entre la valeur moyenne $- 1,96 \times$ écart type des différences et la valeur moyenne $+ 1,96 \times$ écart type des différences). L'étendue de cet intervalle (sa hauteur) renseigne sur l'interchangeabilité des deux méthodes diagnostiques mais non sur leur valeur intrinsèque.

BOÎTE À MOUSTACHES

Box plot

La boîte à moustaches (*whiskers*), appelée aussi diagramme de surface, est une représentation graphique de la distribution des valeurs médianes et de leurs interquartiles (percentiles 25 % et 75 %). La figure 3 représente schématiquement ce type de diagramme, très informatif et utilisé quand la distribution de la variable n'est pas « normale » ou gaussienne.



Figure 3. Représentation d'une « boîte à moustaches » (box plot) montrant la médiane et les différentes valeurs à prendre en compte.

Min : valeur minimale ; Max = valeur maximale ; Q1 = 25^e percentile ; M. = médiane (50^e percentile), Q3 = 75^e percentile.

BONFERRONI (CORRECTION OU TEST DE)

➔ CORRECTION DE BONFERRONI

C

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

CAS-TÉMOIN (ÉTUDE)

Case control study

Il s'agit d'une étude rétrospective dans laquelle un groupe de patients (ayant la maladie ou traité selon la méthode à évaluer) est comparé à un groupe contrôle (n'ayant pas la maladie ou traité selon une méthode de référence). Les sujets des deux groupes doivent être aussi similaires que possible en dehors du critère évalué.

Ce type d'étude constitue à l'origine une méthode d'enquête explicative en épidémiologie : par exemple, comparer deux groupes de sujets exposés ou non à un facteur de risque présumé et analyser la relation entre ce facteur et l'apparition de la maladie (telle que tabac et cancer du poumon). Il s'agit alors d'une enquête explicative. Dans ce cas, c'est l'*odds-ratio* (ou rapport de cotes) qui est calculé pour évaluer le risque.

Ces études sont relativement rapides à réaliser mais comportent des risques de biais tels que des erreurs de type 2. C'est pour cela qu'on considère qu'elles ont une validité moindre que les études de cohorte (voir *infra*).

CHI²

Chi square

Indice statistique permettant d'estimer le degré d'accord entre une distribution théorique et celle obtenue expérimentalement dans une étude. En partant de l'hypothèse nulle (deux méthodes diagnostiques ou thérapeutiques sont similaires), on peut calculer les valeurs attendues dans deux groupes. Le calcul du *Chi²* ou χ^2 consiste à apprécier l'écart entre les valeurs observées et les valeurs attendues selon la formule détaillée dans le tableau II.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Par exemple, si on considère la comparaison entre une méthode diagnostique A et une méthode B, les valeurs observées peuvent être les suivantes :

- diagnostic fait par A = a, et diagnostic fait par B = b ;
- diagnostic non fait par A = c et diagnostic non fait par B = d.

Les valeurs attendues correspondantes sont a', b', c', d'.

La valeur du χ^2 permet ensuite, à l'aide d'une table, de déduire la valeur de p.

La formule du χ^2 est résumée dans le tableau II.

Tableau II. Calcul de χ^2 .

	Méthode		
	A	B	
Diagnostic fait	a	b	a + b
Diagnostic non fait	c	d	c + d
	a + c	b + d	N = a + b + c + d
$\chi^2 = \frac{(a - a')^2}{a'} + \frac{(b - b')^2}{b'} + \frac{(c - c')^2}{c'} + \frac{(d - d')^2}{d'}$			

COCHRAN (TEST Q DE)

➔ TESTS STATISTIQUES

COCHRAN-MANTEL-HAENSZEL (TEST DE)

➔ TESTS STATISTIQUES

COCHRANE (COLLABORATION)

Cochrane Collaboration

La Collaboration Cochrane (www.cochrane.org) est une organisation internationale indépendante à but non lucratif créée en 1993 et portant le nom d'Archie Cochrane, un épidémiologiste britannique. La mission de cette organisation est de fournir une information sur les soins de santé régulièrement mise à jour et disponible à travers le monde. Le principal rôle de la Collaboration Cochrane est de produire des revues systématiques et de promouvoir les meilleures preuves scientifiques sous la forme d'essais cliniques bien conduits. La principale production de la Collaboration Cochrane est la base de données des revues systématiques (*Cochrane Database of Systematic Reviews*) publiée dans la Librairie Cochrane (voir *infra*). Ces revues

systématiques, ou méta-analyses, sont préparées et rédigées par un groupe de revues (*Collaborative Review Group*) selon des normes méthodologiques rigoureuses. Une cinquantaine de groupes forme la Collaboration Cochrane à travers le monde. Chaque groupe s'occupe d'une pathologie ou d'une entité spécifique.

COCHRANE (LIBRAIRIE)

Cochrane Library

C'est une base de données payante (www3.interscience.wiley.com/cgi-bin/mrwhome/) mise à jour tous les trimestres et incluant plusieurs registres :

- le registre des revues systématiques (méta-analyses) faites par (et selon la méthodologie de la Collaboration Cochrane) un groupe Cochrane (*Cochrane Database of Systematic Reviews*) ;
- le registre des résumés de revues systématiques faites en dehors de la Collaboration Cochrane et généralement publiées dans des revues indexées dans *PubMed (Database of Abstracts of Reviews of Effects)* ;
- le registre central des essais randomisés publiés (*Cochrane Central Register of Controlled Trials*, ou CENTRAL) ;
- le registre des revues méthodologiques faites dans le cadre du Cochrane (*Cochrane Database of Methodology Reviews*) ou hors de ce cadre (*Cochrane Methodology Register*).

La base de données contient aussi les données du programme d'évaluation technologique (*Health Technology Assessment Database*, ou HTA) et économique (*NHS Economic Evaluation Database*) de la santé par le ministère de la Santé anglais.

La Cochrane Library est incontournable pour les médecins qui veulent appliquer les principes de la médecine factuelle (*Evidence-Based Medicine*) dans leur pratique quotidienne.

COHORTE (ÉTUDE DE)

Cohort study

Il s'agit d'une étude prospective longitudinale, comparative ou non, d'un groupe de patients. Dans le cas d'une étude non comparative, le groupe de patients est suivi de manière prospective dans le but d'évaluer l'exposition à un risque particulier (par exemple, le risque de cancer du poumon chez les fumeurs). Ce type d'étude nécessite généralement un grand nombre de patients pour estimer de manière fiable le risque associé à l'exposition. Dans les études comparatives, le plus difficile est de définir *a priori* le groupe contrôle. Deux types de groupes contrôles peuvent être constitués : soit un groupe « historique » ayant les mêmes

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a caractéristiques en dehors du facteur évalué (comme l'âge ou le sexe)
b ou apparié sur ces caractéristiques, soit un groupe « contemporain »
c parallèle où les patients sont suivis pendant des mois ou années.
d L'étude randomisée n'est qu'une forme particulière d'étude de
e cohorte avec un groupe parallèle contemporain, mais moins sujette
f à des biais puisque l'allocation est aléatoire grâce à la randomisation.
g Lorsqu'une randomisation n'est pas faisable, les études prospectives
h de cohorte sont considérées comme celles qui sont les moins
i sujettes à biais, surtout lorsque le recrutement est exhaustif.

COMITÉ D'ÉTHIQUE

➔ **CONSENTEMENT ÉCLAIRÉ**

CONCORDANCE (TEST DE)

Concordance or agreement testing

Le test de concordance s'applique lorsque l'on veut savoir si l'interprétation des résultats d'un examen diagnostique (une radiographie, par exemple) ou les réponses de patients à un questionnaire sont concordantes.

➔ **CORRÉLATION**

CONSENTEMENT ÉCLAIRÉ

Informed consent

Le consentement éclairé est l'adhésion librement consentie d'une personne (ou d'un représentant légal) à participer à une étude clinique. La personne doit avoir été clairement informée du but, de la méthode de l'étude, des avantages et inconvénients de toutes les procédures concernées, ainsi que du degré d'incertitude du résultat. Le consentement éclairé est régi par la loi du 9 août 2004, complétée par le décret du 26 avril 2006, qui modifie en profondeur le cadre juridique de la recherche biomédicale. Ces nouvelles règles applicables à l'encadrement de la recherche biomédicale se substituent au dispositif issu de la loi du 20 décembre 1988, dite loi Huriet-Sérusclat. Les détails de cette nouvelle loi sont disponibles sur le site : www.recherche-biomedicale.sante.gouv.fr.

Le consentement éclairé et écrit concerne toute la recherche biomédicale ou toute étude pouvant modifier la prise en charge diagnostique ou thérapeutique habituelle. Cela inclut les essais randomisés et les études non randomisées (études cas-témoins, études de cohorte) avec ou sans bénéfice individuel direct.

CONSORT (CONSOLIDATED STANDARDS OF REPORTING TRIALS)

CONSORT Statement

Le CONSORT (www.consort-statement.org) est une description des essais cliniques randomisés sous la forme d'une liste d'items à laquelle doit satisfaire tout essai soumis pour publication. C'est en quelque sorte un critère de qualité méthodologique. La liste contient 22 items, comme le montrent le tableau III et la figure 4. Toutes les grandes revues médicales ont adopté le CONSORT pour évaluer les essais randomisés soumis pour publication. Généralement l'essai n'est accepté pour publication que s'il satisfait aux critères du CONSORT. Le lecteur peut ainsi juger de la validité interne et externe de l'essai publié.

Tableau III. Liste des items inclus dans le CONSORT.

Section	Numéro	Descripteur	Page où l'item est rapporté
Titre et résumé	1	Décrire l'étude comme étant une étude avec « répartition au hasard », ou « randomisé »	
Introduction :			
• Antécédents	2	Décrire les données scientifiques et fournir le rationnel de l'étude	
Méthode			
• Participants	3	Décrire les critères d'éligibilité et les lieux de l'étude	
• Intervention	4	Fournir les détails précis des interventions réalisées dans chaque groupe	
• Objectifs	5	Énoncer l'hypothèse définie de façon prospective et les objectifs de l'étude	
• Données (critères de jugement)	6	Définir les critères de jugement primaires et secondaires et les méthodes éventuellement utilisées pour améliorer la qualité des mesures et éviter les biais	
• Taille de l'effectif	7	Décrire comment a été calculée la taille de l'effectif, les éventuelles analyses intermédiaires et les règles d'arrêt de l'étude (si nécessaire)	

a
b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

Répartition :			
• Calendrier de répartition	8	Décrire la méthode utilisée pour le flux répartition, y compris les détails sur les restrictions adoptées (par exemple : blocs, stratifications)	
• Dissimulation de la répartition	9	Décrire la méthode utilisée pour dissimuler la répartition (par exemple : enveloppes numérotées, téléphone centralisé)	
• Exécution	10	Indiquer qui a généré la répartition, qui a inscrit les participants, et qui a alloué les participants aux différents groupes	
• Insu	11	Décrire si les participants, les personnes réalisant l'intervention, les évaluateurs des résultats, et les analystes des données n'étaient pas au courant de l'allocation. Si oui, décrire dans quel degré l'exigence est respectée	
• Analyses statistiques	12	Décrire les analyses utilisées pour comparer les groupes concernant les données primaires ainsi que les données supplémentaires, par exemple : analyses de sous-groupes et covariables	
Résultats :			
• Flux des participants	13	Fournir un profil de l'étude (un diagramme est fortement recommandé). Pour chaque groupe, indiquer le nombre de participants qui ont été inclus, qui ont reçu l'intervention allouée, qui ont complété le protocole et qui ont été inclus dans l'analyse des données. Décrire les écarts du protocole ainsi que les raisons	
• Recrutement	14	Donner la période de recrutement et la durée de suivi	
• Données de base	15	Fournir les données démographiques de base et les caractéristiques cliniques pour chaque groupe de patients	
• Effectif analysé	16	Fournir le nombre de participants (dénominateur) dans chaque groupe et indiquer si les analyses ont été faites en intention de traiter. Fournir les chiffres absolus chaque fois que possible (par exemple : 15/30 et non 50%)	

• Résultats et estimations	17	Décrire l'effet estimé de l'intervention sur les mesures des données primaires et secondaires, y compris une estimation par points et une mesure de précision (par exemple : intervalle de confiance) pour chaque groupe	
• Analyses supplémentaires	18	Fournir toute autre analyse effectuée, y compris les analyses de sous-groupes et covariables, indiquant lesquelles étaient prospectives et lesquelles étaient définies a posteriori	
• Effets indésirables	19	Décrire tous les effets indésirables notables ainsi que les effets secondaires pour chaque groupe	
Commentaires :			
• Interprétation	20	Fournir une interprétation des résultats, en tenant compte de l'hypothèse de l'étude, des sources d'imprécision, et des dangers associés à la multiplicité des analyses	
• Généralisabilité	21	Discuter la validité externe des résultats de l'étude	
• Vue d'ensemble	22	Donner une interprétation générale des données à la lumière de l'ensemble des données factuelles disponibles	

(D'après Moher D. Schulz F. K. Altman G. D. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials, *Lancet* 2001, 357 ; 1191-4. Avec la permission d'Elsevier Masson SAS – traduction française de l'auteur K. Slim).

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

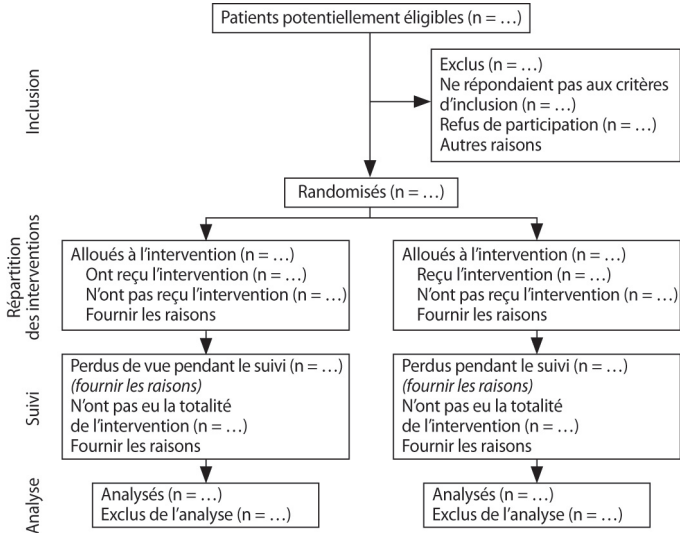


Figure 4. Diagramme des inclusions dans un essai randomisé tel qu'il est exigé dans le CONSORT.

(D'après Moher D. Schulz F. K. Altman G. D. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials, *Lancet* 2001, 357 ; 1191-4. Avec la permission d'Elsevier Masson SAS – traduction française de l'auteur K. Slim).

CORRECTION DITE DE BONFERRONI

Bonferroni correction

Elle consiste à diviser le risque α par le nombre n de comparaisons à tester. Si on compare 3 valeurs de pression artérielle entre elles (1 avec 2, 1 avec 3 et 2 avec 3), le risque α sera de $0,05/3 = 0,016$. Si une valeur de p est inférieure à $0,016$, on écrira alors que la différence est significative au risque α de $0,05$.

Exemple : comparaison de plusieurs traitements

Comparaison de trois agents anesthésiques (B, C et D) à un agent anesthésique A sur l'incidence des vomissements postopératoires : cette incidence est de 40 % avec B, 30 % avec C, et 20 % avec D ($n = 100$ dans chaque groupe). L'analyse statistique réalisée en comparant A versus B, A versus C, A versus D conduit à conclure que seul D est différent de A. Du fait des comparaisons multiples, il faut bien

noter que le risque de première espèce consenti n'est plus de 0,05 mais de 0,11 (3 comparaisons). Si l'on comparait 100 agents anesthésiques à l'agent A, avec un risque $\alpha = 0,05$, on pourrait attendre que 5 d'entre eux soient différents de A, seulement par hasard ! L'analyse statistique doit tenir compte des comparaisons multiples afin de garantir *in fine* un risque global de première espèce de 0,05.

À partir du moment où la comparaison statistique est effectuée sur plusieurs moyennes (plus de 2), il n'est pas possible d'utiliser un test prévu pour comparer deux moyennes une seule fois. Il convient alors d'utiliser une méthode statistique qui garantisse la conservation du risque $\alpha = 0,05$. L'une des méthodes est la correction de Bonferroni.

CORRECTION DE YATES

Yates correction

Lorsqu'on calcule la valeur de p en utilisant une distribution continue du Chi^2 à partir de variables discrètes, la valeur de p peut être sous-estimée. Pour réduire ce risque, certains statisticiens conseillent d'ajuster le calcul du Chi^2 vers le bas (car pour un degré de liberté fixe, p évolue dans le sens contraire du Chi^2).

CORRÉLATION (COEFFICIENT DE)

Coefficient of correlation

La corrélation est une association entre deux séries de données. Elle peut être positive (association dans le même sens) ou négative (dans le sens contraire). Le calcul du coefficient de corrélation tend à montrer dans quelle mesure l'association est linéaire (pouvant être schématisée par une ligne droite).

La figure 5 schématise les différents résultats obtenus après un test de corrélation.

Deux types de coefficients sont généralement calculés : le coefficient de Pearson pour les variables paramétriques (r ou r_{xy} , noté de -1 à $+1$) et le coefficient de Spearman pour les variables non paramétriques (r_s , noté de 0 à 1 , une valeur proche de 1 signifiant une bonne corrélation). Le coefficient de Spearman (corrélation de rang) est utilisé lorsque les variables ne sont pas réparties normalement (distribution non gaussienne) ou lorsqu'une variable (ou deux) est mesurée sur une échelle ordinale (binaire ou plus).

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

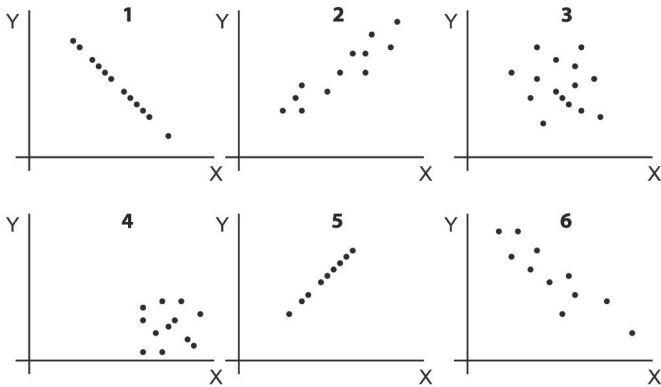


Figure 5. Types de corrélations entre une variable X et une variable Y.
1 = corrélation négative parfaite ; 2 = bonne corrélation positive ; 3 et 4 = absence de corrélation ; 5 = corrélation positive parfaite ; 6 = bonne corrélation négative.

COTES (RAPPORT DES)

Odds ratio

Le rapport des cotes (RC) est un rapport de risque. C'est le rapport entre le risque (cote) de survenue d'une maladie ou d'un événement dans un groupe et celui dans un groupe contrôle. La formule du RC est faite à partir d'un tableau de contingence classique (2×2), comme le montre le tableau IV.

Cela revient à calculer le rapport entre deux cotes dans les groupes comparés. Un RC égal à l'unité signifie un risque identique, un RC > 1 signifie une augmentation du risque dans le groupe évalué, et inversement. Plus le RC est éloigné de 1, plus le risque ou l'effet d'un traitement est important.

Exemple

Deux groupes de 8 patients sont comparés. Dans le groupe A, 2 patients ont présenté l'événement (prenons l'exemple d'une complication chirurgicale), d'où une fréquence de l'événement de $2/8$, soit 0,25. Dans le groupe B, 5 patients ont présenté l'événement, soit une fréquence de $5/8$, soit 0,62. Le calcul du RC est schématisé dans le tableau IV. La cote pour le groupe A est de $0,25/0,75 = 0,33$; la cote pour le groupe B est de $0,62/0,38 = 1,63$. Le rapport de cotes pour la survenue de la complication dans le groupe B est de $1,63/0,33 = 4,94$.

Enfin, lorsque la fréquence de l'événement est faible (inférieure à 20 %), le RC et le risque relatif sont très voisins. Le RC est donc une estimation du risque relatif lorsque la fréquence de l'événement est faible. *A contrario*, en termes d'efficacité d'un traitement, le RC aura tendance à surestimer l'effet du traitement quand le risque de base est élevé (au-dessus de 25 % environ). Dans ce cas, les résultats présentés avec un RC seront plus favorables au traitement que ceux fondés sur le risque relatif, et un RC ne peut être interprété comme un risque relatif.

Tableau IV. Formule du rapport des cotes.

	Sujet malade (ou ayant une complication)	Sujet sain (ou n'ayant pas de complication)
Exposé (ou groupe A)	a	b
Non exposé (ou groupe B)	c	d
RC = $[f_b / (1 - f_b)] / [f_a / (1 - f_a)]$ ou bien RC = $(a/c) / (b/d)$ soit $(a \times d) / (b \times c)$		

La formule du rapport des cotes (RC) est dérivée du tableau de contingence 2×2 où, dans une étude épidémiologique, a représente les personnes exposées au risque et devenues malades, b, les personnes exposées mais saines, c, les personnes non exposées au risque mais malades, et d, les personnes non exposées et saines.

Le même raisonnement est valable dans les cas de comparaisons entre deux traitements (A versus B), où a représente par exemple les patients du groupe A ayant présenté une complication, b, les patients du groupe B ayant présenté la même complication, c, les patients du groupe A n'ayant pas présenté la complication, et d, les patients du groupe B n'ayant pas présenté la complication. Fréquence de la complication dans le groupe A : $f_A = a / (a + c)$; fréquence de la complication dans le groupe B : $f_B = b / (b + d)$.

COURBE D'APPRENTISSAGE

Learning curve

La courbe d'apprentissage, ou courbe d'acquisition technologique, représente la période au-delà de laquelle les effets de l'inexpérience (effets indésirables, complications, imperfections) disparaissent ou sont significativement réduits. Le concept découle d'une idée simple : plus on pratique une intervention, meilleurs sont les résultats. Elle concerne habituellement des domaines techniques comme la chirurgie, la radiologie ou l'endoscopie interventionnelle, ou l'interprétation d'une méthode diagnostique.

La courbe d'apprentissage consiste à recueillir la fréquence de survenue d'un événement donné (morbidité, durée opératoire, durée de séjour) au fil du temps et à relever le nombre de procédures au-delà

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

duquel cette fréquence rejoint une valeur moyenne définie *a priori*. Cette valeur est généralement celle observée soit dans la pratique courante, soit avec une intervention contrôlée. Cette valeur seuil (*cut-off point* ou *threshold point*) correspond au nombre d'interventions nécessaire pour acquérir une expérience. Plus cette valeur seuil est élevée, plus l'intervention est jugée difficile.

Deux modèles théoriques permettent de calculer la courbe d'apprentissage : le modèle de Wright (1936) et le modèle de Crawford (plus largement utilisé). Une analyse détaillée de ces modèles peut être consultée sur le site www.maaw.info/LearningCurveSummary.htm.

Mais la notion de courbe d'apprentissage pose des problèmes méthodologiques : comment fixer la valeur seuil ? La courbe d'apprentissage varie-t-elle selon que l'opérateur est seul ou au sein d'une équipe ? *Quid* des courbes sans valeur seuil claire ?

Elle pose aussi des problèmes éthiques : comment gérer les patients traités pendant la période d'apprentissage ? Faut-il inclure tous les patients pendant cette période dans une étude clinique ? Comment raccourcir la période d'apprentissage ?

Deux exemples de courbes d'apprentissage en chirurgie

En chirurgie colorectale par laparoscopie, une analyse multivariée a suggéré que le nombre nécessaire d'interventions pour dépasser la période d'apprentissage était de 40 interventions (Bennett C. L. *et al. Arch Surg* 1997, 132 : 41-5). *A contrario*, une autre étude a montré que le nombre de fundoplicatures par laparoscopie pour dépasser la période d'apprentissage n'était que de 20 interventions (Watson D. I. *et al. Ann Surg* 1996, 224 : 198-203). Ces résultats suggèrent que l'apprentissage de la chirurgie colorectale par laparoscopie est plus difficile que celui de la fundoplicature par la même voie.

COURBE DE SURVIE

➔ SURVIE

COURBE ROC

Receiver Operating Characteristic curve

Le nom de la courbe ROC vient des militaires anglais qui avaient, pendant la Seconde Guerre mondiale, développé ce principe pour régler leurs radars afin de mieux détecter les missiles allemands. La courbe ROC est un graphique exprimant la capacité d'un test de dépistage ou d'un test diagnostique à faire la distinction entre les sujets sains et les sujets malades. La sensibilité du test (vrais positifs) est indiquée sur l'axe des ordonnées et les faux positifs (1 - spécificité) sur l'axe des abscisses. La valeur optimale se trouve dans

le coin supérieur gauche du graphique, c'est-à-dire le point de la courbe le plus éloigné de la diagonale.

La figure 6 représente un exemple de courbe ROC.

L'aire sous la courbe indique la performance du test ou sa précision. Une aire égale à 1 signifie que le test est parfait, il n'y a ni faux positif ni faux négatif. Plus l'aire sous la courbe est proche de 1, meilleure est la précision du test. Une aire égale à 0,5 signifie que le test est médiocre, il y a autant de vrais que de faux positifs.

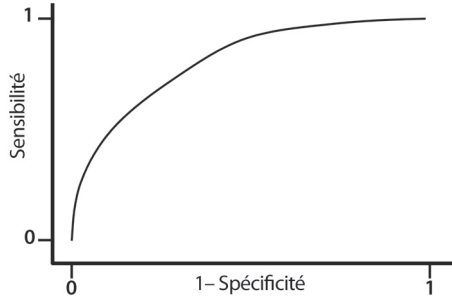


Figure 6. Exemple de courbe ROC.

La valeur optimale est ici de 1, ce qui signifie que le test évalué est performant.

COÛT-BÉNÉFICE (ANALYSE)

Cost-benefit analysis

Les études de coût-bénéfice permettent de déterminer si une nouvelle stratégie de santé dégage un bénéfice net pour la société. Elles comparent le coût au montant que la société peut dégager pour obtenir une unité supplémentaire de soins. Ces analyses, qui englobent tous les coûts (directs et indirects) et les conséquences (exprimées en termes monétaires) de la stratégie en question, renseignent sur la disposition de la société à payer. Les différents types d'évaluation économique sont résumés dans le tableau V.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Tableau V. Différents types d'évaluation économique.

Type	Mesure des coûts	Identification des conséquences	Mesure des conséquences	Critère
Minimisation des coûts	Monétaire	Conséquences identiques pour toutes les options	Aucune	$C_1 - C_2$
Coût-efficacité	Monétaire	Indicateur d'efficacité à dimension unique	Unité physique	$\frac{C_1 - C_2}{E_1 - E_2}$
Coût-utilité	Monétaire	Indicateur d'efficacité à plusieurs dimensions	QALY	$\frac{C_1 - C_2}{QALY_1 - QALY_2}$
Coût-bénéfice	Monétaire	Indicateur d'efficacité à plusieurs dimensions	Monétaire	$(C_1 - B_1) - (C_2 - B_2)$

C = coût ; E = efficacité ; B = bénéfice ; QALY = Quality Adjusted Life Year. Les chiffres 1 et 2 en indice sont en rapport avec les stratégies 1 et 2 comparées dans le cadre des analyses économiques.

(D'après le Guide méthodologique de la HAS – nous remercions la Haute Autorité de Santé de nous avoir autorisés à reproduire ce tableau. Le texte intégral est de plus consultable sur le site www.has-sante.fr, rubrique Publications).

COÛT-EFFICACITÉ (ANALYSE)

Cost-effectiveness analysis

C'est une évaluation médico-économique d'une stratégie de dépistage, de diagnostic ou de traitement d'une affection donnée. Cette évaluation peut entrer dans le cadre d'une décision publique concernant l'adoption ou la régulation d'une innovation biomédicale ou d'une aide à la décision diagnostique ou thérapeutique. Les études de coût-efficacité mettent en relation les coûts comparés avec des résultats en termes d'indicateurs cliniques quantitatifs ou d'objectifs de l'état de santé (décès, survie). Il s'agit donc de rechercher la meilleure stratégie diagnostique ou thérapeutique. Les études coût-efficacité permettent d'apporter une information aux décideurs sur le supplément d'efficacité obtenu au travers d'un supplément de coût.

Un diagramme décisionnel (figure 7) est généralement établi pour résumer les différentes situations rencontrées.

Pour les économistes de la santé, l'analyse coût-efficacité tient compte de l'incertitude des deux dimensions, coût et efficacité, et consiste généralement à appliquer des tests statistiques au ratio (rapport) des différences estimées, d'où un ratio coût-efficacité incrémental.

mental, qui constitue actuellement la forme courante des résultats de l'analyse coût-efficacité (correspondant au coût net d'un changement de stratégie).

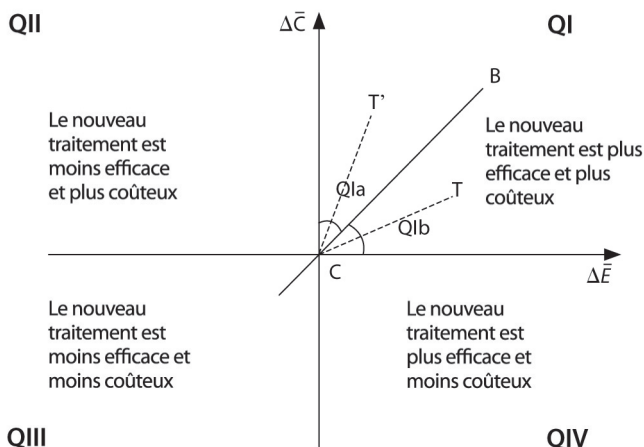


Figure 7. Diagramme de coût-efficacité selon Black.

Comparaison de deux stratégies thérapeutiques : l'axe des ordonnées représente les différences de coût (ΔC) et l'axe des abscisses les différences d'efficacité (ΔE). Les 4 quadrants correspondent aux différentes possibilités : QI, le nouveau traitement est plus efficace et plus coûteux ; QII, le nouveau traitement est moins efficace et plus coûteux ; QIII, le nouveau traitement est moins efficace et moins coûteux ; QIV, le nouveau traitement est plus efficace et moins coûteux. Le quadrant QI est divisé par la ligne CB en deux parties : la partie inférieure, où se trouve le point T , correspond à un traitement recommandé et la partie supérieure, où se trouve le point T' , correspond à un traitement devant faire l'objet d'une discussion de politique de santé. Le quadrant QII correspond au plus mauvais cas de figure en termes de coût-efficacité et le quadrant QIV correspond au meilleur cas de figure.

(D'après Black WC. The CE plane : a graphic representation of cost-effectiveness. *Med Decis Making* 1990 ; 10 : 212-4.)

COÛT-UTILITÉ (ANALYSE)

Cost-utility analysis

C'est une forme particulière d'analyse coût-efficacité. Les résultats médicaux sont exprimés non plus en termes d'indicateurs objectifs

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

(décès, survie) mais en termes d'équivalents d'années de vie gagnées pondérés par la qualité (*Quality adjusted life years*, ou QALY) ou d'autres indicateurs d'utilité. Les mesures ne sont dans ce cas plus seulement quantitatives mais aussi qualitatives. On dispose ainsi d'un critère unique résumant ces deux aspects.

Trois exemples

- En cancérologie, deux stratégies thérapeutiques, la chirurgie et la radiothérapie exclusive, peuvent être comparées en termes de coût-utilité concernant d'une part la survie, et d'autre part la morbidité et l'amélioration (ou l'aggravation) de la qualité.
- En pathologie fonctionnelle, deux stratégies peuvent être comparées : le traitement médical (inhibiteurs de la pompe à protons) et la fondoplicature laparoscopique.
- Dans la lithiase de la voie biliaire principale, deux stratégies : chirurgie laparoscopique en un temps et traitement en deux temps (sphinctérotomie endoscopique puis cholécystectomie laparoscopique), peuvent être comparées en termes de coût-utilité.

COX (MODÈLE DE)

Proportional hazards (Cox) model

Le modèle de Cox obéit aux principes des analyses multivariées. Il est utilisé lorsque la variable analysée est une variable censurée comme la mortalité ou la récurrence. Il s'agit d'un modèle de régression à effet proportionnel, c'est-à-dire estimant l'influence de facteurs de risque sur la probabilité que se produise un événement au cours d'une période de temps donnée, en partant de l'hypothèse que cette influence est constante dans le temps.

La mesure de l'association fournie par le modèle de Cox est un rapport de risque (*risk-ratio*) qui exprime un risque relatif entre la variable analysée et une covariable.

Sur un plan mathématique, le modèle de COX permet d'identifier les facteurs indépendants expliquant le risque de survenue d'un événement qui est lié au temps : $h(t) = h_0(t) \times e^{\beta X}$, h étant le *hazard* des Anglais, c'est-à-dire le risque, X la ou les variables indépendantes, β le coefficient de la régression.

Exemple : la transplantation hépatique est-elle bénéfique aux patients ayant un hépatocarcinome ?

Dans un monde parfait, la réponse viendrait d'un essai randomisé, mais la pratique nous montre qu'un tel essai est très difficile à réaliser. On peut donc appliquer ce modèle à une série de patients transplantés. Le but serait de vérifier si la transplantation hépatique est une variable indépendante influençant la durée de survie. Mais les

résultats de la transplantation dépendent aussi des délais d'attente sur la liste de l'Agence de biomédecine. Cette variable doit être prise en compte pour évaluer la covariable transplantation. C'est le principal intérêt de ce modèle.

➔ **VARIABLE**

CRITÈRE D'INCLUSION

Inclusion criteria

C'est l'ensemble des critères décrivant les conditions auxquelles doivent satisfaire les patients pour être inclus dans une étude. Ils sont déterminés *a priori* avant l'étude et doivent répondre aux objectifs de l'étude. En même temps que les critères d'inclusion, doivent aussi être définis les critères de non-inclusion (les patients ne devant pas être inclus dans l'étude) et les critères d'exclusion (quand un patient déjà inclus doit sortir de l'étude).

Ces critères renseignent sur la validité externe de l'étude, c'est-à-dire le type de patients auxquels doivent être appliquées les conclusions de l'étude.

CRITÈRE DE JUGEMENT

Endpoint – Outcome

Le critère de jugement est la mesure permettant d'évaluer le résultat d'un événement ou d'une intervention thérapeutique. Il peut s'agir d'un critère objectif (décès, glycémie) ou d'un critère subjectif (douleur, qualité de vie). On distingue généralement un (des) critère(s) de jugement principal (aux) ou fort(s), qui est (sont) décidé(s) de manière pragmatique (survie à 5 ans après résection d'un cancer, récurrence après cure de hernie) et un (des) critère(s) secondaire(s). Ces critères principaux et secondaires sont définis *a priori* avant le début de l'étude.

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

d

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

DELPHES (MÉTHODE)

➔ **TECHNIQUE DELPHES**

DICHOTOMIQUE (VARIABLE)

Dichotomous variable

Appelée aussi binaire, la variable dichotomique (à l'inverse de la variable continue) est exprimée sous forme de deux classes ou deux catégories, comme oui/non, présent/absent, vivant/décédé.

➔ **VARIABLE**

DIFFÉRENCE MOYENNE

Mean difference

La différence moyenne peut être standardisée ou pondérée. La différence moyenne pondérée mesure la taille (ampleur) de l'effet pour des variables continues (taille, poids, score de gravité) plutôt que dichotomiques (ou binaires). Les différences moyennes entre les groupes sont pondérées pour prendre en compte des différences en termes de taille de l'échantillon, précision des études. Cette différence moyenne pondérée est une valeur absolue exprimée en unité de mesure du critère de jugement considéré.

À l'opposé, la différence moyenne standardisée prend en compte les différences en termes de méthode de mesure (échelle ou score de douleur). Elle est exprimée sous forme de rapport sans unité de mesure.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

DISCONTINUE OU DISCRÈTE (VARIABLE)

Discrete variable

À la différence des variables continues qui prennent un continuum de valeurs (c'est-à-dire qu'entre deux valeurs possibles, il en existe toujours une troisième), les variables discontinues sont distinctes. Elles sont aussi appelées variables discrètes.

Les histogrammes (et non les courbes) sont la meilleure représentation des variables discontinues.

Exemples

- Dans un jet de dé : on ne peut avoir que 6 valeurs (ou variables) discontinues.
- Nombre de grossesses ou d'enfants.
- Nombre de poussées de diverticulite sigmoïdienne avant chirurgie.

DISCRIMINANTE (ANALYSE)

Discriminant (analysis)

L'analyse discriminante a un objectif différent de celui des autres analyses multivariées. Elle permet, à l'aide de covariables, une discrimination entre deux groupes ou plus, en fonction d'une variable à expliquer. La figure 8 schématise le principe de l'analyse discriminante. Les résultats sont exprimés pour chaque covariable (ou ensemble de covariables) par le calcul du coefficient de Mahalanobis, du p , et du pourcentage de sujets correctement classés (probabilité bayésienne). Le tableau VI présente un exemple d'analyse discriminante des facteurs de mortalité à 1 mois des patients cirrhotiques ayant eu une hémorragie digestive.

Tableau VI. Analyse discriminante des facteurs de mortalité à 1 mois des cirrhotiques ayant eu un épisode d'hémorragie digestive.

<i>Analyse discriminante unidimensionnelle</i>			
<i>Variable</i>	<i>Coefficient de Mahalanobis</i>	<i>p</i>	<i>% de patients bien classés</i>
Ascite	0,364	< 0,01	63
Bilirubinémie	0,303	< 0,01	70
Temps de Quick	0,286	< 0,01	67
Cause de l'hémorragie	0,158	< 0,05	61
Médicaments gastroagressifs	0,154	< 0,05	58
Type de l'hémorragie	0,085	NS	48

Analyse discriminante multidimensionnelle		
Covariables	Coefficient de Mahalanobis	% de patients bien classés
Ascites	0,364	63
Ascite + bilirubinémie	0,587	72
Ascite + bilirubine + cause	0,764	72
Ascite + bilirubine + cause + Quick	0,864	75
17 variables	1,710	79

(D'après Poynard T. *et al. Gastroenterol Clin Biol*, Masson, 1980 ; 4 : 655.)

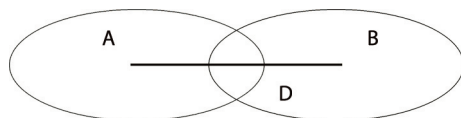


Figure 8. Représentation de l'analyse discriminante de deux groupes A et B. Plus la droite D (joignant les centres des ellipses A et B) est longue, mieux les deux groupes peuvent être discriminés. La valeur D² est appelée coefficient de Mahalanobis (biostatisticien indien). Plus le coefficient est grand, meilleure est la discrimination. Il est possible de calculer pour un test diagnostique les vrais positifs, les faux positifs, les vrais négatifs ou les faux négatifs, donc des valeurs comme la sensibilité, la spécificité, et les valeurs prédictives positive et négative.

DISTRIBUTION NORMALE (GAUSSIENNE)

Normal (Gaussian) distribution

Souvent appelée simplement « courbe en cloche », la courbe suit une loi dite « normale » (loi de Gauss-Laplace). Le plus grand nombre d'individus ou de valeurs est au milieu, et on constate une symétrie par rapport au milieu de la courbe. La moyenne (moyenne arithmétique de l'ensemble des nombres), la médiane (valeur centrale lorsque les données sont ordonnées de la plus basse à la plus haute, la moitié des données se situe au-dessus, tandis que l'autre moitié lui est inférieure) et le mode (valeur la plus fréquente) sont égaux, et les résultats situés à chaque extrémité de la distribution sont les moins fréquents. Par exemple, dans une courbe représentant les résultats d'un test d'intelligence, la plupart des personnes se trouvent au centre ou autour de l'intervalle d'intelligence « moyen », tandis que le nombre de personnes diminue des deux côtés à mesure que les résultats s'éloignent de la moyenne, d'où la forme et le nom de la courbe.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Les figures 9 et 10 montrent deux exemples de courbes gaussiennes.

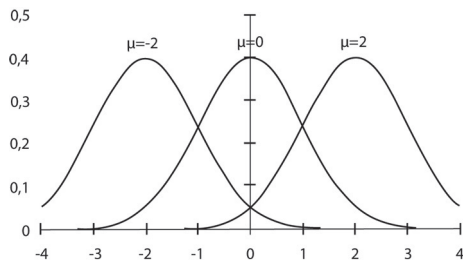


Figure 9. Courbe de Gauss de la loi normale notée $N(\mu, 1)$ pour les valeurs de μ (espérance ou moyenne vraie) de -2 , 0 , et $+2$.

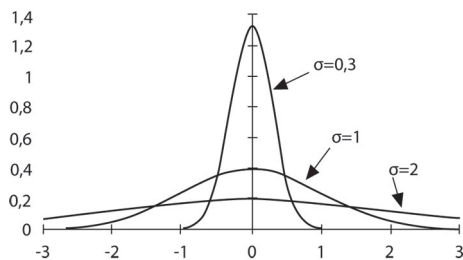


Figure 10. Courbe de Gauss de la loi normale notée $N(0, \sigma^2)$ pour les valeurs de σ (écart type) de $0,3$, 1 , et 2 . La forme de la courbe est ainsi déterminée par la valeur de l'écart type (σ).

DOUBLE AVEUGLE

Double blind

➔ AVEUGLE

e

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

ÉCART TYPE

Standard deviation

L'écart type (comme la variance) est un indicateur de dispersion. L'écart type (ou ET, souvent désigné par la lettre s ou la lettre grecque σ) indique comment en moyenne les valeurs de la variable sont groupées autour de la tendance centrale x (moyenne arithmétique). Un écart type faible signifie que les valeurs sont peu dispersées autour de la moyenne (série homogène) et, inversement, un grand écart type signifie une dispersion importante (série hétérogène). Dans une distribution normale (gaussienne), 95 % des valeurs sont comprises entre $-1,96$ ET (à gauche) et $+1,96$ ET (à droite) de la moyenne.

Quelques règles d'utilisation de l'ET sont à connaître : on n'utilise l'ET que pour mesurer la dispersion autour de la moyenne d'un ensemble de données. L'ET n'est jamais négatif, et est sensible aux valeurs aberrantes. Une seule valeur aberrante peut accroître l'ET et donc déformer l'aspect de la dispersion.

ÉCHELLES

Scales

Les échelles de mesure permettent de transformer une donnée qualitative en donnée quantitative. On peut ainsi établir des scores ou des index. On distingue schématiquement quatre types d'échelles :

- échelle nominale : les nombres sont attribués sans ordre défini ; par exemple, 1 pour les femmes et 2 pour les hommes, ou bien 1 pour vivant et 2 pour décédé. L'échelle nominale donne une information sur l'identité ou l'état. Elle est souvent utilisée pour le recueil de données ;

EFFET HAWTHORNE

Hawthorne effect

On appelle effet Hawthorne l'influence (négative ou plus souvent positive) d'une intervention due au simple fait de participer à une étude de recherche. Le nom de cet effet date des années 1920, où il a été constaté dans une étude sur l'effet de l'éclairage du plan de travail à la *Western Electric Company* de Hawthorne que la productivité était aussi améliorée dans le groupe contrôle (sans éclairage). Cet effet a été constaté en médecine dans beaucoup d'essais en cancérologie ou sur la douleur.

EFFET PLACEBO

Placebo effect

C'est l'effet souvent positif psychologique, physiologique ou psychophysiologique de tout médicament ou de toute intervention à visée thérapeutique, qui est indépendant, ou ne dépend que de façon minime, des effets spécifiques de la prescription ou de l'intervention. Cet effet est attribué à d'autres facteurs comme l'histoire naturelle de la plainte, la relation médecin/patient, l'attente du patient, du médecin ou de l'investigateur. En recherche clinique, les médicaments peuvent être comparés à une substance placebo. Cette comparaison est difficile à envisager en chirurgie pour des raisons éthiques évidentes.

EFFICACITÉ

➔ **AMPLEUR DE L'EFFET**

EFFICIENCE

Efficiency

L'efficacité d'un traitement ou d'une intervention signifie son efficacité par rapport aux moyens matériels et humains mis en œuvre. Le traitement ou l'intervention sont d'autant plus efficaces qu'ils sont efficaces à moindre coût.

➔ **COÛT-EFFICACITÉ, COÛT-BÉNÉFICE**

EMBASE

Embase

C'est une base de données informatique ayant succédé à la base *Excerpta Medica*, éditée par Elsevier, et accessible sur le site www.sciencedirect.com. Embase contient des références de 1974 à nos

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

a jours. Chaque année 450 000 références sont ajoutées et il y a 40 %
b de chevauchement avec Medline. L'accès à Embase n'est pas gratuit
c et nécessite un abonnement. La version actuelle d'Embase est appe-
d lée Scopus (www.scopus.com).

ERREUR DE TYPE 1 OU DE PREMIÈRE ESPÈCE

→ ALPHA (RISQUE)

ERREUR DE TYPE 2 OU DE DEUXIÈME ESPÈCE

→ BÊTA (RISQUE)

ERREUR TYPE

Standard error

L'erreur type (ET) d'un paramètre statistique est l'écart type de la « distribution » de ce paramètre, ce qui est différent de l'écart type « tout court » de ce paramètre. Plus la taille de l'échantillon est grande, plus l'ET est réduite. Si la mesure concerne une moyenne, l'ET est dite erreur autour de la moyenne, ou *standard error of the mean (SEM)*. Cette mesure consiste, en cas d'échantillons multiples, à calculer la moyenne de chaque échantillon. L'ensemble des moyennes forme une distribution normale (ou gaussienne) et l'écart type de cette distribution constitue l'ET de la moyenne ou SEM. Quand on dispose d'un seul échantillon, l'ET de la moyenne est égale à l'ET de l'échantillon divisée par la racine carrée de l'effet.

L'ET sert à calculer l'intervalle de confiance. Par exemple, dans un essai randomisé, l'intervalle de confiance à 95 % est estimé à la différence entre deux valeurs $\pm 1,96$ ET.

ESSAI CLINIQUE

Clinical trial

Un essai ou une étude clinique est généralement réalisé pour tester l'efficacité d'un médicament, d'une stratégie thérapeutique, ou d'une intervention chirurgicale.

On distingue pour les médicaments (mais aussi par extension pour les interventions chirurgicales) 4 phases :

- essai de phase I : après la phase expérimentale sur les animaux, le traitement est testé soit sur des volontaires sains, à la recherche d'éventuels effets indésirables, ou du dosage le moins dangereux du produit, soit sur des patients (par exemple, en cancérologie) pour déterminer la dose maximale tolérée. Le métabolisme du

produit testé est étudié au cours de cette phase. La recherche clinique est faite sans recours à un groupe contrôle ;

- essai de phase II : il consiste à évaluer l'efficacité et la tolérance de différents schémas thérapeutiques (doses, fréquences d'administration, modalités thérapeutiques). Les groupes inclus dans l'essai sont souvent de faible effectif. Il n'y a généralement pas de distribution aléatoire, mais certains essais de phase II peuvent être randomisés ;
- essai de phase III : c'est un essai clinique incluant un grand nombre de patients pour évaluer l'efficacité et l'innocuité du traitement. Cet essai est dans la majorité des cas randomisé et comporte au moins 2 groupes d'évaluation. On parle généralement de groupe traité (patients recevant le traitement testé) et de groupe contrôle (patients recevant un placebo ou un traitement standard de référence) ;
- essai de phase IV : cet essai est réalisé après la mise sur le marché d'un produit pharmaceutique ou la validation d'une intervention chirurgicale. L'efficacité et les effets indésirables sont analysés sur une longue période et un grand nombre de patients (pharmacovigilance, matériovigilance). Il s'agit généralement d'une étude de cohorte sans groupe témoin.

ESSAI CLINIQUE CROISÉ

Cross-over clinical trial

À la différence de l'essai contrôlé randomisé classique où les groupes sont parallèles, l'essai croisé comporte le passage à un moment de l'essai du patient d'un groupe à l'autre. Le patient devient ainsi son propre témoin. La figure 11 montre un exemple d'essai randomisé croisé.

Exemple

Dans un essai français sur la neuromodulation pour incontinence anale, tous les patients subissaient l'implantation d'un stimulateur sous-cutané, et la randomisation (à l'insu du patient) consistait à mettre le stimulateur en position « marche » ou « arrêt ». À un moment de l'étude, les patients du groupe « stimulateur en marche » étaient basculés dans le groupe « arrêt » (à leur insu, pour éliminer un effet placebo) et inversement. L'analyse de l'efficacité de ce système en était donc rendue plus objective.

Le principal avantage des essais croisés est de diviser la taille de l'effectif (nombre de sujets nécessaires) par deux, même si sur le plan statistique les choses sont un peu plus complexes, car pour un critère de jugement quantitatif le nombre de sujets dépend aussi du coefficient de corrélation (pour diviser le nombre de sujets par deux,

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

il faut que le coefficient de corrélation soit proche de zéro). Il est nécessaire malgré tout de vérifier que l'ordre de passage imposé au patient n'a pas influencé le résultat (ANOVA multiple).

Un autre avantage de l'essai croisé est de quantifier l'effet d'un traitement et d'en évaluer l'effet relatif même s'il est jugé moins efficace. Par exemple, en comparant un traitement A à un traitement B dans un essai classique, on trouve que A est supérieur à B. Dans un essai croisé, on peut rechercher la proportion de patients ne répondant pas au traitement A mais pouvant bénéficier du traitement B. Cette approche rappelle une situation clinique habituelle « dans la vie réelle » : quand un traitement est inefficace chez un patient donné, on en essaie un autre, même si on sait qu'il est moins efficace que le précédent.

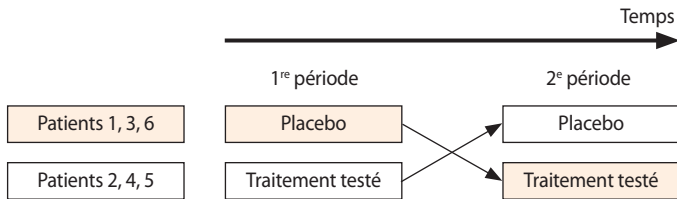


Figure 11. Schéma représentant un exemple d'essai croisé. Les patients sont d'abord randomisés dans un groupe puis basculé dans l'autre groupe (à leur insu) à un moment donné de l'essai.

ESSAI CONTRÔLÉ RANDOMISÉ

Randomized controlled trial (RCT)

À la différence du simple essai clinique contrôlé (comparaison d'un ou plusieurs groupes traités à un ou plusieurs groupes contrôles), l'essai contrôlé randomisé (ECR) est un essai habituellement de phase III qui comporte la réalisation d'un tirage au sort (distribution ou allocation ou affectation aléatoire) pour répartir les patients dans les différents groupes de l'étude. L'ECR constitue la meilleure méthode de recherche clinique pour évaluer l'efficacité d'un traitement ou tester une hypothèse car la distribution aléatoire réduit le risque de biais.

L'ECR peut comparer : un groupe traité avec un groupe témoin recevant un placebo, deux traitements médicaux, un traitement médical et un traitement chirurgical, deux techniques chirurgicales, ou des associations thérapeutiques ou des stratégies thérapeutiques entre elles.

Le protocole de l'ECR est établi de manière prospective, définissant les groupes de l'étude, les critères de jugement, la technique de ran-

domisation (par exemple, issue de tables de nombres au hasard), la stratification, le nombre de sujets nécessaire à inclure. Le *CONSORT Statement* a défini les normes de conduite et de publication des ECR.

ESSAI D'ÉQUIVALENCE ET ESSAI DE NON-INFÉRIORITÉ

Equivalence trial and non-inferiority trial

Un essai clinique d'équivalence a pour objectif de montrer que deux traitements sont équivalents en termes d'efficacité clinique. Dans la majorité des cas, on se « contente » de montrer que le nouveau traitement B n'est pas moins efficace que le traitement de référence A. Il s'agit alors d'essai de non-infériorité. Contrairement aux classiques essais randomisés de supériorité, la méthodologie des essais d'équivalence ou de non-infériorité est encore peu connue. Dans ce type d'essais, les valeurs du risque d'erreur α et β sont inversées.

Les circonstances pouvant justifier la réalisation d'un essai d'équivalence sont résumées dans le tableau VII.

Parmi les difficultés de ces essais, citons-en deux :

- ils demandent l'inclusion d'un grand nombre de patients ;
- surtout, ils nécessitent la définition d'un « seuil (ou zone) d'équivalence », en deçà (ou à l'intérieur) duquel on considère que les traitements évalués sont équivalents. Ce seuil correspond à une différence négligeable sur le plan clinique, habituellement définie de manière empirique par un consensus d'experts.

Dans un essai d'équivalence, le choix du seuil d'équivalence est essentiel car la signification statistique d'un résultat dépend étroitement de la limite d'équivalence choisie. Le test sera significatif avec une limite très tolérante et non significatif avec une limite plus stricte. La figure 12 illustre l'importance du choix du seuil d'équivalence.

Pour illustrer les problèmes posés par les essais de non-infériorité, prenons un exemple et un contre-exemple.

Exemple

Dans l'essai de non-infériorité du COST Study Group comparant la chirurgie conventionnelle à la chirurgie laparoscopique pour cancer colorectal (*N Engl J Med* 2004, 350 : 2050-9), le critère de jugement était le délai de la récurrence. Les auteurs avaient calculé qu'il fallait inclure 1 200 patients. Pour que la voie laparoscopique ne soit pas considérée comme inférieure, la différence du risque de récurrence ne devait pas dépasser 22 %, avec une valeur seuil de p à 0,41.

Contre-exemple

Il s'agit de l'essai de Lacy *et al.* (*Lancet* 2002, 359 : 2224-9). Les auteurs avaient choisi comme seuil d'équivalence une différence du taux de mortalité par cancer à 15 % ! Ce qui leur permettait de

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

n'inclure que près de 200 patients. Mais peut-on considérer une différence de 15 % comme étant cliniquement négligeable en matière de cancérologie ?

Tableau VII. Circonstances pouvant justifier la réalisation d'un essai d'équivalence ou de non-infériorité.

Recherche d'effets secondaires moindres ou moins graves
Recherche d'une utilisation plus aisée : <ul style="list-style-type: none"> • voie d'administration plus simple (par exemple, orale par rapport à intraveineuse, bolus à la place d'une perfusion) • prise quotidienne unique à la place de plusieurs prises • ajustement de dose non nécessaire
Recherche d'un traitement moins invasif : <ul style="list-style-type: none"> • traitement médical plutôt que traitement chirurgical • chirurgie moins délabrante (par exemple, laparoscopie) • radiothérapie moins prolongée
Recherche d'un coût plus faible

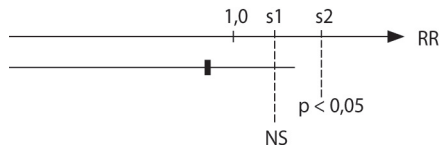


Figure 12. Choix du seuil d'équivalence.

En fonction du seuil d'équivalence choisi, le même résultat peut conduire à :
 – un test de non-infériorité statistiquement significatif (seuil s2), avec un intervalle de confiance en deçà du seuil d'équivalence s1 ;
 – ou un test de non-infériorité non statistiquement significatif (seuil s1), avec un intervalle de confiance dépassant le seuil d'équivalence s2.
 (D'après Cucherat M, www.spc.univ-lyon1.fr.)

ESTIMATION DE KAPLAN-MEIER

Kaplan-Meier estimation

Kaplan et Meier ont développé en 1958 (*J Am Stat Assoc* 1958, 53 : 457-81) une méthode non paramétrique d'estimation de la survie. Cette méthode consiste en une analyse univariée (contrairement au modèle de Cox) prenant en compte le fait que certains patients ne pourront pas être suivis jusqu'à la survenue de l'événement étudié (par exemple, le décès par cancer). La courbe a la forme de marches d'escalier (figure 13),

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

où chaque trait vertical représente la survenue de l'événement étudié (par exemple, le décès). Comme dans une courbe de survie actuarielle (dont la représentation est différente de celle de l'estimation de Kaplan-Meier), il est utile de représenter l'intervalle de confiance à 95 % (pointillés dans la figure 13). Les autres données nécessaires pour l'interprétation des courbes de Kaplan-Meier sont : les caractéristiques de la population étudiée (nombre de patients, âge, stade du cancer, facteurs pronostiques, interventions), le nombre initial de patients étudiés par groupe, la définition de l'événement étudié, le suivi médian, le résultat des tests statistiques, les variables tronquées, les données manquantes et leur distribution dans chaque groupe. L'estimation de Kaplan-Meier trouve ses limites quand le taux de données manquantes dépasse 20 % dans les groupes étudiés, quand les effectifs étudiés sont réduits ou quand les décès postopératoires immédiats sont exclus de l'analyse.

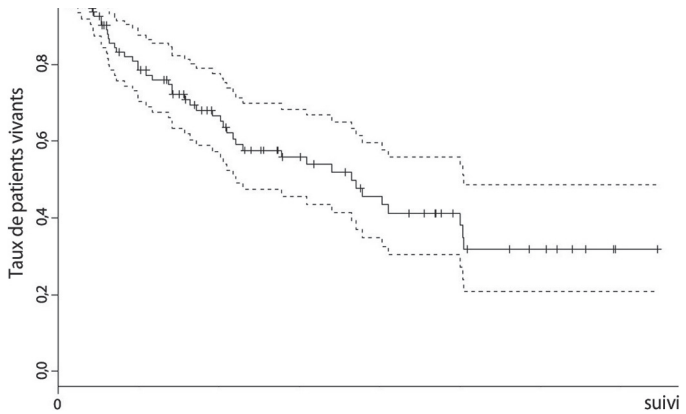


Figure 13. Courbe d'estimation de Kaplan-Meier. Les traits verticaux des marches d'escalier de la courbe pleine représentent les événements étudiés et les courbes en pointillés représentent les intervalles de confiance à 95 %.

ÉTUDE MULTICENTRIQUE

Multicentre study

Il s'agit d'une étude comparative ou non incluant plusieurs centres d'investigation. Ces études posent des problèmes d'ordre technique et d'ordre éthique :

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

- problèmes techniques : nécessité d'une parfaite coordination et harmonisation entre les équipes, notamment en termes de respect du protocole lorsqu'il s'agit d'un essai prospectif (par exemple, un contrôlé randomisé), et définition d'un centre de monitoring ;
- problèmes éthiques de partage de responsabilité et de propriété intellectuelle : les différents participants à une étude multicentrique ne doivent pas se sentir frustrés au profit de l'investigateur principal, ainsi des règles précises de publication doivent être définies *a priori*. Les noms de tous les médecins ayant participé à l'étude doivent être indiqués dans la publication définitive.

ÉTUDE PROSPECTIVE

Prospective study

Une étude est dite prospective lorsque le protocole de l'étude est établi et rédigé avant le début de l'étude (et donc des inclusions). Les données à analyser sont définies *a priori* et leur recueil est fait afin de réduire les données manquantes, sources de biais. Une étude prospective peut être comparative ou non.

ÉTUDE RÉTROSPECTIVE

Retrospective study

Une étude rétrospective comporte l'analyse de données survenues dans le passé. Ce type d'étude est associé à un faible niveau de preuves, du fait :

- des biais relatifs aux données manquantes ;
- de l'absence d'équivalence entre les groupes étudiés (en cas d'étude comparative historique) ;
- du manque de critères pertinents d'analyse (non recueillis à l'époque mais devenus indispensables depuis).

ÉTUDE TRANSVERSALE

Cross-sectional study

L'étude transversale est une étude épidémiologique descriptive, réalisée sur un échantillon donné, à un moment donné, afin de collecter des informations telles que des facteurs de risque ou d'autres données. Ces études sont menées, par exemple, pour calculer la prévalence d'une maladie ou recenser les infections nosocomiales.

EXACTITUDE DIAGNOSTIQUE

➔ **PRÉCISION DIAGNOSTIQUE**

f

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

FACTEUR DE CONFUSION

Confounding factor

Un facteur de confusion est un facteur qui comporte une association avec le facteur évalué en cas d'étude épidémiologique, ou pouvant introduire un biais dans le résultat en cas d'étude thérapeutique.

L'exemple le plus classique en chirurgie est la comparaison de la morbidité de deux interventions alors que les patients ne sont pas similaires, avec un pronostic ou un risque opératoire plus mauvais dans un groupe. Cette situation est théoriquement éliminée dans les essais randomisés où la répartition se fait de manière équilibrée (effet du hasard), mais l'exemple 3 ci-après montre qu'elle est possible.

On peut contrôler l'effet d'un facteur de confusion en réalisant soit un appariement des populations étudiées, soit une analyse multivariée à la recherche de facteurs indépendants.

Exemple 1

Une association entre la consommation de café et un risque accru d'infarctus du myocarde est décrite. En fait, il est démontré que les fumeurs boivent plus de café que les non-fumeurs et que ce sont les premiers qui ont un risque accru d'avoir un infarctus du myocarde. Le café n'augmente pas à lui seul le risque, c'est son association avec le tabac (facteur de confusion) qui est à l'origine de l'augmentation du risque.

Exemple 2

Dans une étude comparative entre la colectomie gauche par laparotomie et la colectomie gauche par cœlioscopie, on trouve plus de sténoses anastomotiques après cœlioscopie. En fait les chirurgiens ont réalisé plus d'anastomoses mécaniques par cœlioscopie. La technique d'anastomose est donc un facteur de confusion qui aurait pu faire croire que c'est la voie cœlioscopique (simple voie d'abord) qui causait plus de sténoses.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Exemple 3

Dans une étude randomisée comparant la cœlioscopie à la laparotomie dans le cancer du côlon (Lacy *et al. Lancet* 2002, 359 : 2224), la survie était meilleure après cœlioscopie, ce qui peut suggérer que la cœlioscopie est supérieure à la laparotomie. Or il s'avère que le taux de patients ayant eu une chimiothérapie adjuvante est plus élevé dans le groupe cœlioscopie. La chimiothérapie peut donc être un facteur de confusion à l'origine de la meilleure survie.

FACTEUR D'IMPACT

➔ SCIENCE CITATION INDEX

FACTUELLE (MÉDECINE)

Evidence-Based Medicine (EBM)

La traduction française du terme EBM est multiple : médecine factuelle, médecine fondée sur les preuves, médecine fondée sur les meilleures preuves, médecine fondée sur des preuves probantes, médecine fondée sur des faits démontrés, médecine fondée sur des données probantes, mais pour des raisons de simplicité on peut utiliser le terme « médecine factuelle ».

La médecine (ou la chirurgie) factuelle est « l'intégration consciencieuse, explicite et judicieuse des meilleures preuves scientifiques dans la prise en charge d'un patient donné » (Sackett DL *et al. Evidence-Based Medicine : what it is and what it isn't. Br Med J* 1996, 312 : 71-2). Ce nouveau paradigme a été initié au Canada (Université de McMaster, 1985) sous l'impulsion d'une équipe de chercheurs. Il est succinctement résumé dans le tableau VIII. Ce paradigme a évolué depuis 20 ans, il a été récemment mis à jour en tenant compte du contexte clinique et du choix du patient (figure 14).

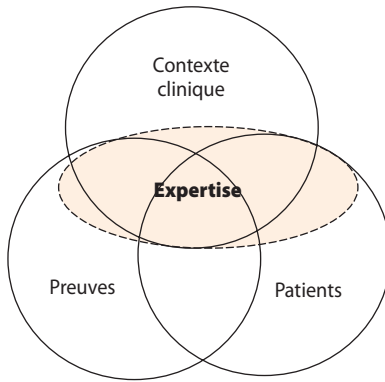


Figure 14. Concept actuel de la médecine factuelle. Ce concept tient compte des circonstances cliniques de la prise en charge des données factuelles de la littérature et du choix du patient, l'expertise clinique étant au centre de ce processus pour prendre en compte ces trois composantes. (D'après Haynes R. B. et al. *Br Med J* 2002 ; 324 : 1350.)

Tableau VIII. Étapes de l'approche factuelle dans la prise en charge d'un patient.

1. Poser une question clinique précise (d'ordre thérapeutique, diagnostique ou pronostique)	Analyse critique de la littérature
2. Rechercher la réponse à cette question dans la littérature (sources primaires et sources secondaires)	
3. Apprécier la qualité des études retrouvées et le niveau de preuves associé à la réponse	
4. Évaluer la validité de la réponse en tenant compte du contexte clinique	Application des données au cas particulier d'un patient
5. Appliquer cette réponse dans la pratique quotidienne avec le consentement éclairé du patient concerné	

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

FAUX NÉGATIF

False negative

Il s'agit d'un indice informationnel sur la valeur diagnostique d'un test ou d'un signe clinique. On parle de faux négatif quand le test est négatif ou le signe clinique absent alors que le patient est atteint de la maladie. Les faux négatifs doivent être pris en compte dans le calcul de la prévalence d'une maladie. Ils sont pris en compte (dans le dénominateur) pour calculer la sensibilité et la valeur prédictive négative (d'un test ou d'un signe clinique), c'est-à-dire que plus les faux négatifs sont nombreux, plus la sensibilité et la valeur prédictive négative sont faibles.

Par extension, on peut qualifier un essai thérapeutique de faux négatif quand il comporte une erreur de type 2 ou β (concluant à tort à l'équivalence de deux traitements).

FAUX POSITIF

False positive

Il s'agit d'un indice informationnel sur la valeur diagnostique d'un test ou d'un signe clinique. On parle de faux positif quand le test est positif ou le signe clinique présent alors que le patient n'est pas atteint de la maladie. Les faux positifs sont pris en compte (dans le dénominateur) pour calculer la spécificité et la valeur prédictive positive (d'un test ou d'un signe clinique), c'est-à-dire que plus les faux positifs sont nombreux, plus la spécificité et la valeur prédictive positive sont faibles.

Par extension, on peut qualifier un essai thérapeutique de faux positif quand il comporte une erreur de type 1 ou α (concluant à tort à la différence de deux traitements).

FIABILITÉ

Reliability

Qualité d'un instrument de mesure dont les résultats, obtenus d'essais indépendants et dans des conditions identiques, demeurent stables.

Les statisticiens préfèrent généralement le terme « fidélité » au terme « fiabilité », qui renseigne plutôt sur l'aptitude d'un dispositif à fonctionner sans défaillance pendant une période déterminée. Le terme « reproductibilité » est aussi utilisé comme synonyme de fiabilité.

FISHER (TEST EXACT DE)

➤ **TESTS STATISTIQUES**

FOREST PLOT OU GRAPHIQUE EN FORÊT (OU EN ARBRE)

Forest plot

Il s'agit de la représentation typique d'une méta-analyse. La première publication de *forest plot* date de 1978 (Freiman J. A. *et al. N Engl J Med* 1978, 299 : 690-4), mais cette étude ne comportait pas de méta-analyse. La forme actuelle du *forest plot* revient à S. Evans (communication personnelle ; Lewis S., Clarke M. *Br Med J* 2001, 322 : 1479-80), qui a imaginé en 1983 une forme modifiée de la boîte à moustaches (*box plot*). La première méta-analyse comportant la forme actuelle du *forest plot* date de 1998. Le logo de la Collaboration Cochrane a d'ailleurs repris le *forest plot* d'une méta-analyse en obstétrique (figure 15). Chaque segment horizontal représente une étude incluse dans la méta-analyse, avec le rapport de cotes ou le risque relatif et l'intervalle de confiance (généralement à 95 %) du critère de jugement considéré. La ligne verticale représente l'unité correspondant à l'absence de différence statistique. Le rapport de cotes ou le risque relatif global est situé en bas du diagramme, généralement sous la forme d'un losange (ou diamant), avec son intervalle de confiance. La valeur et la taille du losange renseignent sur l'ampleur de l'effet et sur sa signification. Cette représentation (figure 16) permet aussi d'observer le degré d'hétérogénéité des études incluses dans la méta-analyse.



Figure 15. Logo de la Collaboration Cochrane.

On y voit le forest plot d'une méta-analyse en obstétrique : effet sur la mortalité néonatale des corticostéroïdes en cas de risque d'accouchement prématuré ; les corticostéroïdes diminuent significativement le taux de mortalité. (Avec la permission de la Cochrane Collaboration.)

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

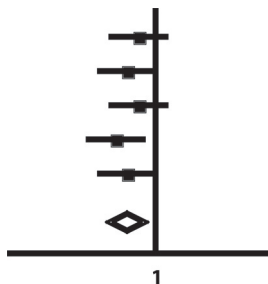


Figure 16. Forest plot simplifié.

Les carrés noirs correspondent au rapport de cotes ou au risque relatif du critère de jugement considéré dans les études incluses dans la méta-analyse. La largeur des lignes horizontales témoigne de l'intervalle de confiance, généralement (mais pas toujours) à 95 %. Le losange correspond au rapport de cotes ou au risque relatif global de la méta-analyse et la largeur du losange à l'intervalle de confiance. Enfin, la ligne verticale correspond à l'unité. Chaque fois qu'un segment horizontal ou le losange croise la ligne verticale (ou l'unité), cela signifie la non-significativité statistique du résultat. Les rapports de cotes ou les risques relatifs supérieurs à 1 témoignent d'un risque supérieur dans le groupe traité par rapport au groupe contrôle.

FRIEDMAN (TEST DE)

➔ TESTS STATISTIQUES

FUNNEL PLOT OU GRAPHIQUE EN ENTONNOIR

Funnel plot

C'est un graphique représentant la dispersion des études incluses dans une méta-analyse à la recherche de biais de publication ; par exemple : études non publiées car négatives (pas de différence retrouvée) ou études de faible effectif non retrouvées par la stratégie de recherche. Chaque point représente une étude et est disposé en fonction de la taille de l'étude et de la taille de l'effet. Idéalement, la disposition des points se présente comme un entonnoir (*funnel*) symétrique (figure 17). En cas de présentation asymétrique (figure 18), on peut évoquer un biais de publication dans la méta-analyse. À l'extrême, en cas d'important biais de publication, la méta-analyse peut n'inclure que les études dites positives, favorisant « artificiellement » le traitement testé.

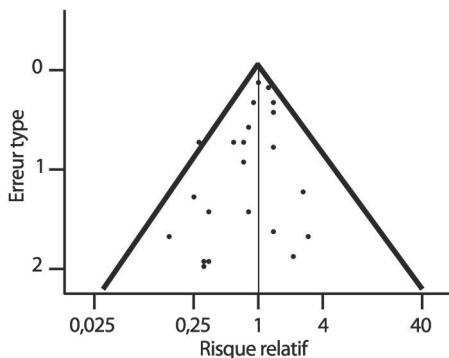


Figure 17. Funnel plot symétrique suggérant l'absence de biais de publication.
 (Avec la permission de la Cochrane Collaboration : www.cochrane.org/resources/handbook/Handbook4.2.6Sep2006.pdf.)

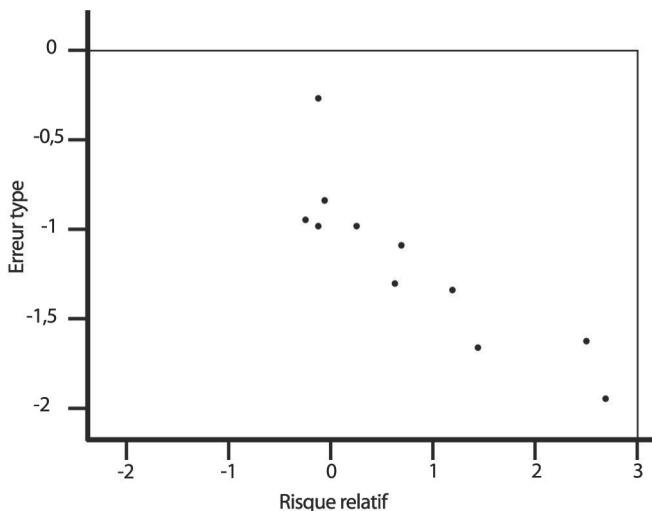


Figure 18. Funnel plot asymétrique suggérant un biais de publication.
 On note un vide dans la partie gauche du graphique signifiant que les études dites négatives ne sont pas incluses dans la méta-analyse.
 (Avec la permission de la Cochrane Collaboration : www.cochrane.org/resources/handbook/Handbook4.2.6Sep2006.pdf.)

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

g

a
b
c
d
e
f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

GOLD STANDARD OU ÉTALON OR OU RÉFÉRENCE (DE)

Gold standard

C'est un test diagnostique ou une méthode thérapeutique ou un instrument de mesure généralement considéré (sur des preuves scientifiques ou de manière consensuelle) comme la meilleure méthode disponible. Dans les essais contrôlés (randomisés ou non), cette méthode de référence est utilisée dans le groupe contrôle pour évaluer de nouvelles méthodes dans le groupe testé.

GRADES DE RECOMMANDATION

Grades of recommendation

Chaque recommandation de pratique clinique factuelle élaborée par une société savante ou un groupe de travail doit comporter une gradation en fonction du niveau de preuves scientifique auquel est associée la recommandation. Il s'agit d'un moyen simple permettant au lecteur d'évaluer la force de la recommandation pour sa pratique clinique. À la différence des niveaux de preuves (utilisant des chiffres), la gradation utilise des lettres (A, B, C, voire D). Plusieurs types de gradation de recommandations ont été publiées, citons celle du *Centre for Evidence-Based Medicine* (CEBM) d'Oxford : le grade A correspond aux niveaux de preuves 1a à 1c, le grade B correspond aux niveaux 2a à 3b, le grade C correspond au niveau 4, et le grade D au niveau 5.

Une gradation plus simplifiée est utilisée par la Haute Autorité de Santé (HAS) en France (www.has-sante.fr/portail/upload/docs/application/pdf/analiterat.pdf) (tableau IX).

➔ **NIVEAUX DE PREUVES**

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Tableau IX. Grades de recommandations, en fonction des niveaux de preuves, utilisés par la Haute Autorité de Santé.

Niveau de preuve scientifique fourni par la littérature	Grade des recommandations
Niveau 1 <ul style="list-style-type: none"> • Essais comparatifs randomisés de forte puissance • Méta-analyse d'essais comparatifs randomisés • Analyse de décision fondée sur des études bien menées 	A Preuve scientifique établie
Niveau 2 <ul style="list-style-type: none"> • Essais comparatifs randomisés de faible puissance 	B Présomption scientifique
Niveau 3 <ul style="list-style-type: none"> • Études comparatives contemporaines non randomisées bien menées • Études de cohorte 	C Faible niveau de preuve scientifique
Niveau 4 <ul style="list-style-type: none"> • Études cas-témoins : essais contrôlés présentant des biais 	
Niveau 5 <ul style="list-style-type: none"> • Études rétrospectives • Séries de cas • Toute étude fortement biaisée 	

(D'après www.has-santé.fr)

GRISE (LITTÉRATURE)

Grey literature

Il s'agit de l'ensemble des données scientifiques (cliniques ou expérimentales) contenues dans des sources difficilement accessibles (contrairement aux bases de données classiques comme par exemple le Medline ou la Cochrane Library). Ce sont, entre autres, les communications orales aux congrès, les thèses non publiées, les dossiers de recherches des laboratoires pharmaceutiques ou fabricants de matériel chirurgical sans caractère commercial. Même si l'avènement d'Internet a rendu l'accès à ces données possible, la recherche de ces données constitue la partie la plus difficile dans la réalisation d'une méta-analyse. Mais la prise en compte de ces données est essentielle si on veut que la méta-analyse soit exhaustive et qu'elle ne comporte pas une exagération de l'ampleur de l'effet thérapeutique. Sinon toute méta-analyse n'est qu'un reflet partiel de la vie réelle.

Exemple

L'évaluation de 135 méta-analyses (McAuley L. *et al.* Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses ? *Lancet* 2000, 356 : 1228-31), choisies au hasard, a montré qu'un tiers comportait des données de la littérature grise et que les études publiées comportaient un effet thérapeutique plus important de 15 % (*Odds Ratio* 1,15 [1,04-1,28]). La prise en compte des seuls articles originaux montrait une différence encore plus significative. Cette étude suggère que l'absence de prise en compte de la littérature grise et l'inclusion des seules études publiées dans les bases de données classiques surestiment l'ampleur de l'effet thérapeutique et donc de l'efficacité des interventions.

GRUPE CONTRÔLE/TÉMOIN

➔ ESSAI CONTRÔLÉ RANDOMISÉ

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

h

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

HAWTHORNE

➔ EFFET HAWTHORNE

HÉTÉROGÉNÉITÉ

Heterogeneity

Les études sont dites hétérogènes lorsqu'elles divergent dans la population étudiée, dans les protocoles utilisés ou dans la méthode d'analyse des résultats. Ainsi, on parle d'hétérogénéité clinique (patients différents, maladies de gravité différente, critères de jugement différents), d'hétérogénéité méthodologique (qualité différente), d'hétérogénéité statistique, à l'origine d'une variabilité de l'effet thérapeutique en fonction des études. Dans les méta-analyses de plusieurs études, la recherche de l'hétérogénéité statistique est une étape essentielle. Elle peut être faite par l'analyse critique des études incluses, la présentation des études sous forme du graphique en arbre (*forest plot*), ou des tests statistiques comme le Q-test, généralement indiqué en bas du *forest plot* pour chaque critère de jugement. À l'extrême, une importante hétérogénéité peut faire surseoir à la réalisation d'une vraie méta-analyse, celle-ci devenant alors une simple revue systématique.

HOMOGÉNÉITÉ

Homogeneity

C'est la similitude entre les études. C'est la situation idéale pour une méta-analyse.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

HYPOTHÈSE NULLE

Null hypothesis

L'hypothèse nulle est le point de départ de tout test statistique permettant de calculer la probabilité que la différence entre deux résultats soit due au hasard. L'hypothèse nulle veut dire que la différence est due au hasard.

Le rejet de l'hypothèse nulle au décours d'un test statistique implique qu'on considère que la différence n'est pas due au hasard mais à l'efficacité réelle d'un traitement. Les tests statistiques permettent de calculer la valeur de p . Le seuil habituel de p pour parler d'une différence « statistiquement significative » est « arbitrairement » fixé à 5 %, c'est-à-dire qu'on a 5 % de risque de rejeter à tort l'hypothèse nulle.

Le seuil classique de 5 % peut varier selon la pertinence clinique de la différence recherchée. Dans certains cas, on fixe la valeur seuil de signification statistique (plus contraignante) à 1 % car on peut considérer qu'un risque de 5 % d'erreur dans le rejet erroné de l'hypothèse nulle est trop élevé.

Exemple

Supposons qu'il existe environ 400 spécialités différentes dans le Vidal® et que chacune n'a été évaluée que par un seul essai thérapeutique. Avec un risque alpha de 5 %, 20 de ces produits seraient présents à tort dans le Vidal® car on aurait rejeté à tort l'hypothèse nulle (Cucherat M. www.spc.univ-lyon1.fr).

Quand on rejette à tort l'hypothèse nulle, on fait une erreur α (ou de type 1) ; quand on accepte à tort l'hypothèse nulle, on fait une erreur β (ou de type 2).

i

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

INCIDENCE

Incidence

Il s'agit d'une mesure descriptive en épidémiologie correspondant au nombre de nouveaux cas recensés pendant une période de temps donnée. On parle généralement d'incidence annuelle. L'incidence peut être exprimée en cas par nombre d'habitants ou par année. Ainsi, on parle :

- d'incidence cumulée, qui est la proportion de population qui développe la maladie sur une période donnée : c'est le nombre de nouveaux cas sur le nombre de personnes composant la population. Par exemple, en France et sur 1 an, 40 nouveaux cas de cancer du côlon chez l'homme sont observés dans une population de 100 000 habitants : l'incidence cumulée est de 40/100 000 ; chez la femme, l'incidence cumulée est de 25/100 000 (Launoy G. *et al. Gastroenterol Clin Biol* 1992, 16 : 633-8) ;
- de taux d'incidence, qui est calculé en divisant le nombre de nouveaux cas au cours d'une période donnée par le nombre de sujets d'une population « à risque ». Par exemple, on parle d'un taux d'incidence de l'infection à VIH de 5 % lorsque 5 nouveaux cas sont diagnostiqués chaque année dans une population de 100 personnes à risque telles que les homosexuels ou les toxicomanes). Le taux d'incidence peut être aussi calculé en années-personnes : par exemple, si 1 000 personnes à risque sont suivies pendant 2 ans et que 26 sont devenues séropositives au VIH pendant cette période, le taux d'incidence est de 1,3 pour 100 années-personnes.

Enfin, contrairement à la prévalence (nombre total de cas), l'incidence permet de suivre l'évolution de la fréquence d'une affection dans le temps. Pour reprendre l'exemple de l'infection à VIH, à la fin des années 1990, l'incidence cumulée baissait grâce aux mesures de prévention, alors que la prévalence continuait à augmenter en rai-

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

son de la longue durée d'incubation et de l'efficacité des traitements qui prolongeaient la vie des malades.

INSU

→ AVEUGLE

INTENTION DE TRAITER (ANALYSE EN)

Intention to treat analysis

En comparant les résultats de deux groupes, seule est considérée la première allocation. C'est-à-dire que même si le patient finit par avoir le traitement alternatif, les événements ultérieurs seront mis sur le compte du traitement qui lui a été alloué initialement. L'analyse en intention de traiter est un critère de bonne qualité des études.

Exemple

Dans une étude comparant la cholécystectomie par coelioscopie et la cholécystectomie par laparotomie, les patients ayant fait l'objet d'une conversion en laparotomie (pour des difficultés opératoires) seront toujours évalués dans le groupe coelioscopie. Les résultats seront rapportés à la coelioscopie.

INTERVALLE DE CONFIANCE

Confidence interval

Donnée « essentielle » pour interpréter les résultats d'une étude clinique ou d'une méta-analyse, l'intervalle de confiance (IC) est une fourchette dans laquelle se situe la valeur réelle. En général, cette fourchette (champ) est calculée afin de se donner 95 chances sur 100 que la réalité se situe entre les valeurs limites (marges de fiabilité ou limites de confiance). On pourrait aussi se fixer, selon la pertinence clinique, des limites moins exigeantes ou, au contraire, plus exigeantes. Dans le domaine des essais thérapeutiques, l'IC permet de visualiser avec précision l'ampleur de l'effet du traitement. Il est plus intéressant que les valeurs de p , qui sont interprétées de manière schématiquement binaire comme « significatives » ou « non significatives ».

L'IC peut être plus ou moins étroit en fonction de la variabilité (exprimée par l'écart type) et de la taille de l'effectif. Plus l'effectif est important, plus l'IC est étroit et proche de la réalité.

Dans un essai thérapeutique ou une méta-analyse, un IC incluant la valeur de non-signification statistique (0 pour la réduction du risque relatif ou absolu et 1 pour le rapport de cotes, ou *odds ratio*, ou

le risque relatif) suggère l'absence d'effet thérapeutique. *A contrario*, un IC n'incluant pas cette valeur suggère un effet thérapeutique. Le principal avantage de l'IC (par rapport à la simple valeur de *p*) est de donner une idée sur l'ampleur de cet effet.

La figure 19 montre un exemple de réduction du risque de mortalité à - 20 % (- 35%— 5 %).

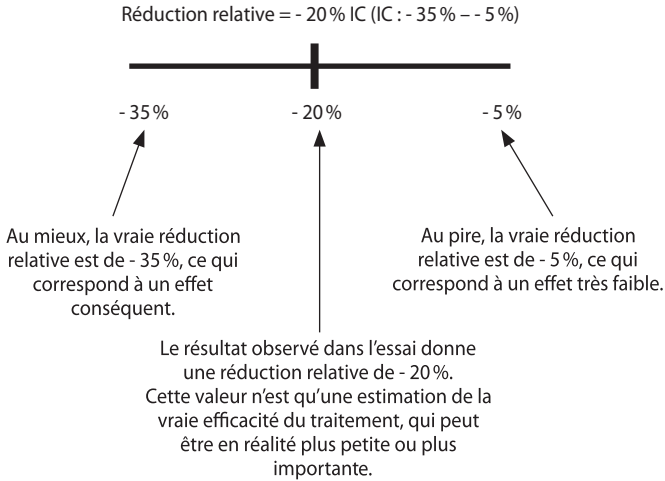


Figure 19. Comment interpréter un intervalle de confiance : ici, dans le cadre d'un calcul de réduction de risque de mortalité.

INTERVALLE INTERQUARTILE

Interquartile interval (or range)

L'intervalle interquartile est l'étendue de la distribution sur laquelle se trouvent concentrée la moitié des éléments dont les valeurs les moins différentes de la médiane. On exclut alors de la distribution 25 % des valeurs les plus faibles et 25 % des valeurs les plus fortes. L'intervalle interquartile est représenté dans le graphique en moustache (*box plot*) sous la forme d'un rectangle.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

k

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

KAPLAN-MEIER

➔ ESTIMATION DE KAPLAN MEIER

KAPPA DE COHEN

Kappa value

Le coefficient kappa (κ) de Cohen évalue la concordance entre deux ou plusieurs observateurs (ou examinateurs), appelée variation interobservateurs, ou entre différentes observations d'une même personne, appelée variation intraobservateur. On calcule la concordance observée (O) et la concordance attendue (A) : la valeur κ est égale à $O - A$ divisé par $1 - A$.

Une valeur κ de 1 correspond à une concordance parfaite, une valeur κ de 0 signifie que la concordance peut être due au hasard, une valeur κ comprise entre 0,4 et 0,6 représente une concordance modérée, une valeur κ de 0,6 à 0,8 représente une bonne concordance, et une valeur κ de plus de 0,8 représente une excellente concordance.

Enfin, dans l'évaluation d'une concordance, les discordances peuvent ne pas avoir le même poids. On parle alors de coefficient κ pondéré. Par exemple, lors de l'évaluation de la concordance entre deux radiologues, conclure sur une tomodensitométrie à une lésion bénigne alors qu'elle est maligne peut être considéré comme plus grave que de ne pas reconnaître un épanchement sous-hépatique. Il convient alors de pondérer le coefficient κ en donnant par exemple comme poids 0 pour les concordances parfaites, 1 pour la discordance considérée comme la moins grave et 3 pour la discordance la plus grave.

KRUSKAL-WALLIS (TEST DE)

➔ TESTS STATISTIQUES

m

MANN-WHITNEY (TEST U DE)

➔ **TESTS STATISTIQUES**

MANTEL-HAENSZEL (TEST DE)

➔ **TESTS STATISTIQUES**

MCNEMAR (TEST DE)

➔ **TESTS STATISTIQUES**

MÉDIANE

Median

La médiane est la valeur centrale (quand les valeurs sont classées par ordre croissant) qui sépare les deux moitiés d'un effectif : 50 % des valeurs sont au-dessus et 50 % sont en dessous. En cas de distribution normale ou gaussienne, la médiane est le plus souvent identique à la moyenne. La médiane peut être estimée sans attendre que les événements se soient produits, comme dans le cadre de la cancérologie où l'on parle de médiane de survie.

MEDLINE

Medline

C'est une base de données électronique produite depuis 1966 par projet du *National Center for Biotechnology Information (NCBI)* de la *National Library of Medicine (NLM)* américaine. Cette base de données est gratuite et accessible sur le site www.pubmed.gov. La base de données contient plus de 15 millions de références venant de 4 332 revues indexées (dont 86 revues françaises) en 2005. C'est une

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

source fondamentale pour la recherche de publications médicales. Elle est la plus largement utilisée du fait de sa gratuité.

MESH

Medical Subject Heading

Le *Medical Subject Heading* (MeSH) est un thesaurus, c'est-à-dire une liste normalisée et structurée de mots acceptés à l'indexation (les descripteurs) et d'équivalents. Ce thesaurus est mis à jour annuellement et traduit en plusieurs langues, dont le français. La maintenance de la version française est faite par l'INSERM. Le tableau X résume les 15 chapitres du MeSH.

Le MeSH contient :

- une liste alphabétique comportant une notice pour chaque descripteur ;
- une arborescence des termes classés par ordre hiérarchique ;
- un index alphabétique listant les mots qui figurent dans les descripteurs.

Contrairement à la recherche par mots libres, la recherche effectuée par mots clés MeSH donne un résultat plus ciblé. Il est aussi admis que les documents non indexés avec les mots clés MeSH sont les moins pertinents. Il est donc conseillé d'utiliser les MeSH dans la recherche sur Medline. D'où la nécessité de consulter le MeSH database (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=mesh), sauf si l'on est certain que le terme recherché ne souffre d'aucune synonymie, ce qui est rare. La recherche par mots libres est soit trop imprécise, soit insuffisante (Clarke M, Greaves L, James S. MeSH terms must be used in Medline searches. *Br Med J* 1997, 314 : 1203). Il convient de toujours associer les deux types de recherche.

Exemple

Si on recherche les articles publiés dans Medline (en mai 2006) sur la laparoscopie dans les cancers coliques, une recherche libre [« colon cancer laparoscopie »] retrouve 604 articles alors qu'une recherche plus ciblée par MeSH [« Colonic Neoplasms/surgery » (MeSH) AND « Laparoscopy/methods » (MeSH) OR « Laparoscopy/surgery » (MeSH)] retrouve 150 articles.

Tableau X. Chapitres des MeSH disponibles dans Medline.

A. Anatomie (<i>Anatomy</i>)
B. Organismes (<i>Organisms</i>)
C. Maladies (<i>Diseases</i>)
D. Produits chimiques, biologiques et pharmaceutiques (<i>Chemicals and Drugs</i>)
E. Équipements et techniques analytiques, diagnostiques et thérapeutiques (<i>Analytical, Diagnosis and Therapeutic Techniques and Equipment</i>)
F. Psychiatrie et psychologie (<i>Psychiatry and Psychology</i>)
G. Sciences biologiques (<i>Biological Sciences</i>)
H. Sciences physiques (<i>Physical Sciences</i>)
I. Anthropologie, enseignement, sociologie et phénomènes sociaux (<i>Anthropology, Education, Sociology and Social Phenomena</i>)
J. Technologie, aliments et boissons (<i>Technology and Food and Beverages</i>)
K. Arts et sciences humaines (<i>Humanities</i>)
L. Sciences information (<i>Information Science</i>)
M. Individus (<i>Persons</i>)
N. Santé (<i>Health Care</i>)
Z. Emplacements géographiques (<i>Geographic Locations</i>)

MÉTA-ANALYSE

Meta-analysis

La méta-analyse consiste à rassembler les données issues d'études comparables (études randomisées voire études non randomisées) et à les analyser au moyen d'outils statistiques adaptés. Elle regroupe les études pertinentes qui essaient de répondre à une question précise de manière critique, reproductible et quantitative. Cette synthèse produit un gain de puissance statistique dans la recherche de l'effet d'un traitement, une précision optimale dans l'estimation de la taille de l'effet et, en cas de résultats apparemment discordants, d'obtenir une vue globale de la situation.

La méta-analyse quantitative utilise soit les données individuelles (retour à la source ou aux cahiers de recueil des données pour chaque patient), soit les données publiées (le cas le plus fréquent), soit de manière prospective les données d'essais en cours.

Il s'agit d'une démarche plus que d'une simple méthode de recherche clinique (tableau XI). La méta-analyse répond au nombre croissant des connaissances sur lesquelles doivent se fonder les décisions médicales et les choix thérapeutiques selon le concept de la médecine factuelle (*Evidence-Based Medicine*), et constitue dans ce contexte le niveau de preuve le plus élevé.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

La méta-analyse n'est applicable que si différentes études utilisent des méthodes similaires et fournissent des données quantitatives exploitables.

À côté de la méta-analyse quantitative, une autre technique de méta-analyse consiste à combiner les résultats disponibles d'études hétérogènes en les interprétant mais sans prétendre fournir des résultats précis et reproductibles : c'est la méta-analyse qualitative (qui devrait dans l'avenir devenir la forme commune de mise au point sur un sujet donné).

Tableau XI. Étapes et avantages de la méta-analyse.

<p><i>Principales étapes</i></p> <ul style="list-style-type: none"> • Collecte exhaustive des données disponibles • Sélection non biaisée des études correspondant à l'objectif et aux critères de sélection préétablis • Constitution de la base de données • Analyse et quantification de l'effet • Interprétation
<p><i>Avantages</i></p> <ul style="list-style-type: none"> • Augmenter la puissance des tests de l'hypothèse que les essais cherchaient à vérifier et la précision de l'estimation de l'intensité (ou la taille) de l'effet • Lever le doute en cas de résultats discordants dans les essais cliniques • Réaliser des analyses de sous-groupes et obtenir des informations qui ne peuvent pas être tirées de chaque étude prise isolément • Constaté l'absence de données fiables sur un sujet précis et amener à la réalisation de nouveaux essais

MODÈLE D'EFFET ALÉATOIRE

Random effect model

Dans une méta-analyse, c'est le degré d'hétérogénéité qui oriente le choix du modèle d'effet d'analyse. Le modèle d'effet aléatoire est une méthode de méta-analyse développée par DerSimonian et Laird (*Meta-analysis in clinical trials. Control Clin Trials* 1986, 7 : 177-88) et utilisée en cas d'hétérogénéité statistique entre les études. C'est un modèle statistique tenant compte du fait que les effets divergents observés dans les études peuvent être dus au hasard mais aussi à des variations réelles entre les études. L'hypothèse d'un modèle d'effet aléatoire est qu'il existe une « population » d'effets éventuels avec une répartition précise autour d'un effet global moyen.

MODÈLE D'EFFET FIXE

Fixed effect model

Le modèle d'effet fixe suppose, à la différence du modèle d'effet aléatoire, qu'il n'y a qu'une seule valeur de l'effet, donc l'absence de toute hétérogénéité des essais. Toute variation observée serait alors liée au hasard. Ce modèle est appliqué dans les méta-analyses lorsque le test d'hétérogénéité n'est pas significatif.

MOOSE (META-ANALYSIS OF OBSERVATIONAL STUDIES IN EPIDEMIOLOGY)

MOOSE

Le MOOSE (*JAMA* 2000, 283 : 2008-12) a été développé selon le même concept que le QUOROM. Il permet d'évaluer la qualité des articles rapportant les résultats de méta-analyses d'études d'observation en épidémiologie.

➔ **QUOROM**

MOYENNE

Mean

C'est la somme des valeurs recueillies divisée par le nombre de sujets. En cas de distribution normale ou gaussienne, la moyenne est identique à la médiane. En cas de faible effectif, la moyenne peut être prise à défaut car il suffit d'un seul événement extrême pour que la moyenne soit décalée de la réalité ; la médiane est alors préférable.

Exemple

Dans le calcul de la durée d'hospitalisation moyenne, si l'effectif est faible (par exemple, une dizaine de patients), un seul patient hospitalisé pendant une longue période (pour une raison donnée, comme par exemple une complication après une intervention chirurgicale) suffit à faire croire à une moyenne élevée. Ainsi, pour 10 patients hospitalisés respectivement 3, 5, 5, 3, 4, 95, 6, 9, 3, 6 jours, la moyenne est de 13,9 jours (tirée vers le haut par le seul patient étant resté hospitalisé pendant 95 jours). La médiane de 5 jours est plus proche de la réalité. Il est de toute manière nécessaire, quelle que soit la formulation du résultat (et pour une interprétation fiable), de donner les deux résultats et d'associer l'écart type à la moyenne et les extrêmes à la médiane.

MULTIVARIÉE

➔ **RÉGRESSION LINÉAIRE MULTIPLE, RÉGRESSION LOGISTIQUE, COX (MODÈLE DE)**

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

n

NIVEAUX DE PREUVES

Levels of evidence

Il s'agit d'un classement hiérarchique de la force des preuves scientifiques associées à chaque affirmation ou recommandation clinique. Le niveau de preuves dépend du type d'études ayant servi à proposer un traitement plutôt qu'un autre, retenir un moyen diagnostique, ou affirmer un pronostic. Plusieurs échelles de niveaux de preuves ont été publiées ou proposées. Elles découlent toutes du même principe : le meilleur niveau de preuves est fourni par les méta-analyses d'essais randomisés de bonne qualité, et le plus bas niveau de preuves est l'avis d'experts sans données scientifiques. L'échelle la plus complète est celle proposée par le *Centre for Evidence-Based Medicine* (CEBM) d'Oxford (mai 2001). Elle est résumée dans le tableau XII. D'autres classifications plus succinctes de niveaux de preuves ont été développées : citons à titre d'exemple celle de la Haute Autorité de Santé (HAS) en France (www.has-sante.fr/portail/upload/docs/application/pdf/analiterat.pdf).

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Tableau XII. Niveaux de preuves selon le CEBM.

Niveau	Traitement/Prévention Étiologie/Risques	Pronostic	Diagnostic	Diagnostic différentiel Étude de prévalence des symptômes	Analyses économiques Analyses de décisions
1a	RM (avec homogénéité) d'ECR	RM (avec homogénéité) de cohortes liminaires RPC validée dans des populations différentes	RM (avec homogénéité) d'études diagnostiques de niveau 1b réalisées dans plusieurs centres cliniques différents	RM (avec homogénéité) d'études de cohortes prospectives	RM (avec homogénéité) d'études économiques de niveau 1
1b	ECR individuel (avec intervalle de confiance étroit)	Étude de cohortes limitaires* individuelle avec un taux de suivi = 80 % RPC validée dans 1 population	Étude de cohortes de validation avec de bons standards de référence RPC évaluée dans 1 centre clinique	Étude de cohortes prospective avec un bon taux de suivi	Analyse fondée sur des coûts ou des alternatives cliniquement sensibles, revue(s) méthodique(s) des preuves et inclusion d'analyses de sensibilité multidirectionnelles
1c	Série de cas de type « tous ou aucun »	Série de cas de type « tous ou aucun »	SpPaf et SnNel absolus	Série de cas de type « tous ou aucun »	Analyses de « meilleure valeur » ou de « pire valeur » absolue
2a	RM (avec homogénéité) d'études de cohortes	RM (avec homogénéité) soit d'études de cohortes rétrospectives, soit d'ECR avec des groupes témoins non traités	RM (avec homogénéité) d'études diagnostiques de niveau > 2	RM (avec homogénéité) d'études de niveau 2b ou meilleur	RM (avec homogénéité) d'études économiques de niveau > 2
2b	Étude de cohortes isolée (y compris ECR de faible qualité, par exemple avec un suivi < 80 %)	Étude de cohortes rétrospective ou ECR avec suivi de témoins non traités Derivation d'une RPC ou validation uniquement sur un échantillon partagé	Étude de cohortes exploratoire avec de bons standards de référence RPC après dérivation ou validation uniquement sur un échantillon partagé ou bases de données	Étude de cohortes rétrospective ou mauvais taux de suivi	Analyse fondée sur des coûts ou des alternatives cliniquement sensibles, revue(s) limitée(s) des preuves ou études isolées et inclusion d'analyses de sensibilité multidirectionnelles

2c	Études de « morbi-mortalité »** Enquêtes épidémiologiques	Étude de « morbi-mortalité »	RM (avec homogénéité) d'études de niveau 3b ou meilleur	Études épidémiologiques	Audit ou études de « morbi-mortalité »
3a	RM (avec homogénéité) d'études cas-témoins	RM (avec homogénéité) d'études de niveau 3b ou meilleur	Étude de cohortes non consécutive ou population très limitée	RM (avec homogénéité) d'études de niveau 3b ou meilleur	RM (avec homogénéité) d'études de niveau 3b ou meilleur
3b	Étude cas-témoins individuelle	Étude de sujets non consécutifs ou sans application homogène du/des standards de référence	Étude cas-témoins, standard de référence mauvais ou non indépendant	Opinion d'expert sans évaluation critique explicite, ou fondée sur la physiologie, des travaux expérimentaux ou des « principes physiopathologiques primaires »	Analyse fondée sur des alternatives ou des coûts limités, des estimations de mauvaise qualité des données, mais avec des analyses de sensibilité incorporant des variations cliniquement sensibles
4	Série de cas (et études de cohortes et cas-témoins de mauvaise qualité)	Série de cas (et études de cohortes pronostiques de mauvaise qualité)	Opinion d'expert sans évaluation critique explicite, ou fondée sur la physiologie, des travaux expérimentaux ou des « principes physiopathologiques primaires »	Série de cas ou standards de référence périmés	Analyse sans analyse de sensibilité
5	Opinion d'expert sans évaluation critique explicite, ou fondée sur la physiologie, des travaux expérimentaux ou des « principes physiopathologiques primaires »	Opinion d'expert sans évaluation critique explicite, ou fondée sur la physiologie, des travaux expérimentaux ou des « principes physiopathologiques primaires »	Opinion d'expert sans évaluation critique explicite, ou fondée sur la physiologie, des travaux expérimentaux ou des « principes physiopathologiques primaires »	Opinion d'expert sans évaluation critique explicite, ou fondée sur la physiologie, des travaux expérimentaux ou des « principes physiopathologiques primaires »	Opinion d'expert sans évaluation critique explicite, ou fondée sur la physiologie, des travaux expérimentaux ou des « principes physiopathologiques primaires »

RM : revue méthodique de la littérature ; ECR : essai comparatif randomisé ; RPC : règle de prédiction clinique ; Splaf : Spécificité si élevée qu'un résultat Positif affirme le diagnostic ; SnNel : Sensibilité si élevée qu'un résultat Négatif élimine le diagnostic.

* Cohorte de patients sélectionnés à un stade précoce de la maladie, indemnes du critère/complication (outcome) d'intérêt.

** Études de morbi-mortalité : traduction proposée de outcomes research ; cependant, n'étant pas structurées selon le plan expérimental d'un essai comparatif randomisé, ces études ne permettent pas de déduire de conclusions solides en termes d'efficacité.

a b c d e f g h i j k l m n o p q r s t u v w x y z

a **NOMBRE DE SUJETS À TRAITER**

b *Number needed to treat (NNT)*

c Le nombre de sujets à traiter (NST) est le nombre de patients qu'il faut traiter pour obtenir un résultat favorable supplémentaire. Ce nombre est calculé par la formule $1/RAR$ (RAR : réduction absolue du risque) avec son intervalle de confiance. La RAR est la différence arithmétique absolue entre le taux d'événements défavorables dans le groupe testé et le taux dans le groupe contrôle.

d **Exemple**

e Dans une méta-analyse sur l'antibioprophylaxie en chirurgie herniaire (Sanchez-Manuel F. J., Seco-Gil J.-L. Antibiotic prophylaxis for hernia repair. *Cochrane Database Syst Rev* 2003 (2) : CD003769), le taux global d'infections était de 3,08 % dans le groupe traité et de 4,69 % dans le groupe contrôle (OR 0,61 [0,32-1,17]). Le NST pour éviter une infection (durant les 30 premiers jours postopératoires) est de près de 60 patients (IC95 % = 25-∞).

k Cette mesure, souvent utilisée dans les méta-analyses, permet d'avoir une représentation plus simple que le rapport de cotes ou le risque relatif. C'est sans doute la mesure la plus facile à comprendre pour un non-initié, et pour convaincre un patient de l'utilité d'un traitement.

m Néanmoins, la présentation du NST, qui ne tient compte que du risque absolu, pose deux problèmes :

- n
- o les NST tirés d'une méta-analyse supposent que les différences de risque sont constantes dans les différents essais alors qu'il peut exister des différences de risque de base ;
 - p l'application du NST à un sujet donné nécessite un ajustement en fonction de son risque de base, d'où l'utilisation de formules plus complexes.

r La formule pour calculer le NST chez un patient donné (tenant compte du risque d'événement selon les circonstances cliniques) est :

s

$$t \quad NST = \frac{1 - [TAE \times (1 - RC)]}{(1 - TAE) \times TAE \times (1 - RC)}$$

u TAE = taux attendu de l'événement (critère de jugement) pris en compte pour calculer le NST ; RC = rapport des cotes (ou *odds ratio*). Ou bien :

v

$$w \quad NST = \frac{1}{TAE \times RRR}$$

x TAE = taux attendu de l'événement (critère de jugement) ; RRR = réduction du risque relatif.

NOMBRE NÉCESSAIRE POUR NUIRE

Number needed to harm (NNH)

Le nombre nécessaire pour nuire (NNN) est le nombre de patients à traiter pour observer un événement « négatif » (effet indésirable d'un traitement, complication chirurgicale, décès). Ce nombre est calculé par la formule $1/AAR$ (AAR : augmentation absolue du risque) avec son intervalle de confiance. De la même manière que le NST pour l'efficacité d'un traitement, le NNN renseigne sur l'innocuité du traitement. Les deux mesures permettent au clinicien et au patient de peser le pour et le contre d'une option thérapeutique donnée et choisir de manière éclairée.

Exemple

Dans une méta-analyse sur le traitement du cancer inopérable de l'œsophage (Wong R., Malthaner R. Combined chemotherapy and radiotherapy (without surgery) compared with radiotherapy alone in localized carcinoma of the esophagus. *Cochrane Database Syst Rev* 2006 ; (1) : CD002092), la radiochimiothérapie est associée à une RAR de 12 % (IC95 % : 3-22 %) des récurrences locales ; le NST pour éviter une récurrence locale est de 9 (pour un taux de récurrence locale de 68 % après radiothérapie seule). Mais cela est associé à une augmentation significative de toxicité sévère : le NNN est alors de 6. Les auteurs de la méta-analyse concluaient que la radiochimiothérapie peut remplacer dans ce contexte la radiothérapie seule, mais à la condition de prévenir les patients du risque de toxicité.

NON-INCLUSION

➔ **CRITÈRES D'INCLUSION**

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

p

PERDU DE VUE

→ BIAIS

PERTINENCE CLINIQUE

Clinical relevance

La pertinence clinique permet de s'assurer que le résultat de l'essai correspond à un effet suffisamment important pour appliquer ses conclusions à la pratique. L'estimation de la taille de l'effet doit être précisée pour évaluer son intérêt en pratique. Cette évaluation repose sur l'emploi des indices d'efficacité (risque relatif, NST) et des intervalles de confiance. Les patients de l'essai doivent être représentatifs des patients vus en pratique médicale courante afin d'assurer la validité externe de l'essai (extrapolation des conclusions).

Exemples

Dans les essais comparant la laparoscopie à la laparotomie en chirurgie colorectale, l'évaluation de la douleur postopératoire (échelle visuelle analogique, consommation de morphine) a montré une différence statistiquement significative en faveur de la laparoscopie. Mais dans certains essais, cette différence était de 10 mg de morphine de moins par jour. Malgré la significativité statistique, la pertinence clinique de cette différence est discutable.

De la même manière, la comparaison de la laparoscopie à l'incision de Mac Burney pour appendicectomie a fait conclure aux auteurs d'une méta-analyse (Sauerland S, Lefering R, Neugebauer EA. *Laparoscopic versus open surgery for suspected appendicitis. Cochrane Database Syst Rev* 2004 ; (4) : CD001546) que certains avantages de la laparoscopie avaient une pertinence clinique limitée.

Enfin, une méta-analyse sur l'antibioprophylaxie avant sphinctérotomie endoscopique (Harris A, Chan AC, Torres-Viera C, Hammett R, Carr-Locke D. *Meta-analysis of antibiotic prophylaxis in endoscopic*

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

retrograde cholangiopancreatography (ERCP). *Endoscopy* 1999, 31 : 718-24.) n'a pas montré d'avantage cliniquement pertinent pour l'antibioprophylaxie en termes de réduction des bactériémies.

PETO (MÉTHODE DE)

Peto Method

La méthode de Peto est appliquée dans les méta-analyses pour calculer les rapports de cotes selon un modèle à effet fixe, c'est-à-dire en l'absence d'hétérogénéité.

PICO

PICO

PICO est l'abréviation de *Patient Intervention Control* (contrôle) *Outcome* (critère de jugement). Il s'agit, dans le cadre de la médecine factuelle (*Evidence-Based Medicine*), d'une manière structurée (et facile à retenir) de poser une question de pratique clinique sur la base d'un scénario donné.

Il s'agit de la première étape de la pratique de la médecine factuelle : poser une question précise. La signification des différents critères du PICO est résumée dans le tableau XIII. La question peut être d'ordre diagnostique, étiologique, thérapeutique ou pronostique.

Récemment, un service d'interrogation de Medline utilisant le principe du PICO a été rendu disponible à l'adresse : <http://askmedline.nlm.nih.gov/ask/pico.php>. Ainsi, plutôt que de faire une recherche par mots libres ou par MeSH, on peut retrouver la liste des données factuelles à travers une question PICO.

Exemple 1

Par quelle voie opérer une cholécystite aiguë (Brugère C, Slim K. *EBM* J 2005, 41 : 2) :

- P : sujets ayant une cholécystite aiguë lithiasique vus en urgence ;
- I : cholécystectomie par laparoscopie dès que le diagnostic est posé ;
- C : cholécystectomie par laparotomie, traitement médical puis cholécystectomie à froid ;
- O : faisabilité de la cholécystectomie par laparoscopie et morbidité postopératoire.

Exemple 2

- P : chez les patients adultes ayant une sinusite aiguë.
- I : traitement de 3 jours à base de triméthoprim-sulfaméthoxazole.
- C : traitement de 10 jours à base de triméthoprim-sulfaméthoxazole.
- O : moins d'effets secondaires à meilleur coût.

Autres exemples de PICO

- Chez un enfant asthmatique, est-ce que les corticoïdes en spray sont à l'origine de plus de retard de croissance que le traitement par bêta-bloquants ?
- Quelle est la durée moyenne de survie après le début d'une insuffisance cardiaque congestive, et quels sont les signes cliniques (s'ils existent) permettant d'identifier les patients pouvant survivre plus ou moins que la moyenne ?

Tableau XIII. Critères du PICO.

Critère P	Caractéristiques du patient (âge, sexe...) et/ou le problème qu'il pose (diagnostic...)
Critère I	Intervention évaluée
Critère C	Comparaison par rapport à une autre intervention (si elle est appropriée)
Critère O	Issue clinique recherchée, critères de jugement et résultats

PLACEBO

➔ EFFET PLACEBO

PONDÉRATION

Weighting

La pondération est l'attribution à chacun des éléments servant à élaborer une moyenne, un indice, un score, d'un coefficient qui exprime son importance relative. La pondération statistique peut concerner par exemple des facteurs prédictifs d'un score, ou des études incluses dans une méta-analyse.

Exemple 1

Dans le cadre de la prédiction de la présence d'une lithiase de la voie biliaire principale au cours d'une cholécystectomie, le score de Huguier (Huguier M. *et al. Surg Gynecol Obstet* 1991, 172 : 470-4) a permis, grâce à une analyse multivariée (régression logistique), d'attribuer à chaque variable un facteur de pondération. La formule est $0,03 \times \text{âge}$ (en années) :

- + 2,2 si voie biliaire principale ≥ 12 mm ;
- + 1,5 si calcul vésiculaire < 10 mm ;
- + 0,7 si antécédents de coliques hépatiques ;
- + 0,8 si cholécystite aiguë.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Exemple 2

Dans le score P-POSSUM (prédiction de la mortalité après chirurgie colorectale : Prytherch D. *et al. Br J Surg* 1998, 85 : 1217-20), l'analyse multivariée a permis d'attribuer au score physiologique (12 items, dont l'âge, la pression artérielle, la biologie) et au score opératoire (6 items, dont les difficultés opératoires, la pathologie, les pertes sanguines) des facteurs de pondération.

La formule du P-POSSUM (R étant le risque de mortalité) est : $\text{Log} [R/(1-R)] = -9,065 + (0,1692 \times \text{score physiologique}) + (0,1550 \times \text{score opératoire})$.

PRÉCISION DIAGNOSTIQUE

Diagnostic accuracy

La précision diagnostique d'un test représente ses performances en termes de sensibilité, spécificité et valeurs prédictives positives et négatives (termes qui seront définis plus loin). En termes statistiques, il n'y a pas de valeur chiffrée de la précision diagnostique. La précision diagnostique d'un test est le degré auquel le résultat du test est un fidèle reflet de la maladie. La figure 20 schématise les caractéristiques d'un test ayant une précision diagnostique idéale et la figure 21 celles d'un test imparfait (le cas le plus fréquent). La précision d'un test peut être représentée par une courbe ROC, avec en ordonnées la « sensibilité » et en abscisse « 1 – spécificité ». Plus l'aire sous la courbe est proche de 1, plus la précision diagnostique est bonne.

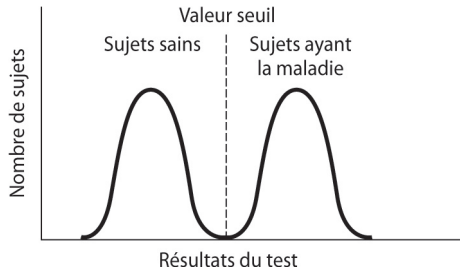


Figure 20. Le test idéal.

Les distributions des résultats du test dans une population de malades et de sujets sains ne se chevauchent pas. Le test a une sensibilité et une spécificité de 100 % si on suppose que la valeur seuil correspond à la ligne verticale en pointillés. Il n'y a ni faux négatifs ni faux positifs.

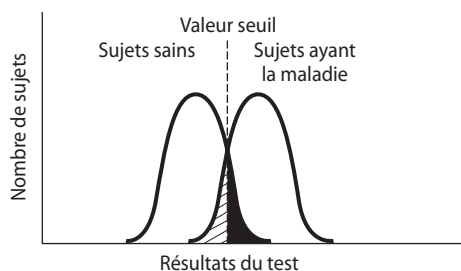


Figure 21. Un test imparfait.

Ici, les distributions des résultats du test dans une population de malades et de sujets sains se chevauchent. La sensibilité et la spécificité sont donc inférieures à 100 %. On peut déplacer la ligne en pointillés correspondant à la valeur seuil, pour modifier la sensibilité et la spécificité du test. Si la valeur seuil est déplacée vers la droite, on améliore la spécificité en réduisant les faux positifs (en noir), mais la sensibilité est diminuée d'autant du fait d'une augmentation des faux négatifs (en hachuré). L'inverse est vrai si on déplace la valeur seuil vers la gauche.

PRÉDICTIONNE (VALEUR)

Predictive value

Les valeurs prédictives positives (VPP) ou négatives (VPN) renseignent sur la probabilité qu'un patient donné (ayant un signe ou un test diagnostique positif) d'avoir (VPP) ou de ne pas avoir (VPN) la maladie.

La VPP est le rapport entre le nombre de vrais positifs (les vrais malades) et l'ensemble des sujets ayant un signe ou un test positif.

La VPN est le rapport entre les vrais négatifs (les vrais sujets sains) et l'ensemble des sujets n'ayant pas un signe ou ayant un test négatif.

Le tableau XIV représente le mode de calcul des valeurs prédictives.

Tableau XIV. Calcul des valeurs prédictives.

	Population de malades	Population de sujets sains	Total
Test (ou signe) positif	a	b	a + b
Test (ou signe) négatif	c	d	c + d
Total	a + c	b + d	

Dans ce tableau, a correspond aux vrais positifs, b aux faux positifs, c aux faux négatifs, d aux vrais négatifs. La VPP est donc égale à : $a/(a + b)$ et la VPN à : $d/(c + d)$.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

PRÉVALENCE

Prevalence

C'est le nombre de malades (nouveaux et anciens) dans une population à un instant donné (prévalence ponctuelle ou *point prevalence*). La prévalence est habituellement exprimée sous forme de pourcentage ou de rapport d'un nombre de malades sur le nombre total de la population examinée ou générale (par exemple : 1/1 000, 1/100 000) ; on parle alors de chiffre de prévalence. Quand on observe le nombre de malades sur une période donnée, on parle de prévalence de période, comme par exemple dans le cadre du sida.

PRÉVENTION

Prevention

La prévention est l'ensemble des mesures prises pour éviter qu'une maladie survienne et qu'elle se complique. On parle de prévention primaire, secondaire et tertiaire. Pour illustrer ces trois types de prévention, nous prendrons l'exemple des affections cardiovasculaires :

- prévention primaire : toutes les interventions visant les facteurs étiologiques de la maladie des sujets qui n'en ont encore jamais présenté de manifestation. Selon l'OMS, c'est « l'ensemble des moyens mis en œuvre pour empêcher l'apparition d'un trouble, d'une pathologie ou d'un symptôme ». Par exemple, en tenant compte des facteurs de risque de maladie coronarienne, la prévention primaire est fondée sur : l'arrêt du tabac, le traitement des hypercholestérolémies, ou un régime alimentaire méditerranéen ;
- prévention secondaire : toutes les actions visant à influencer favorablement sur l'évolution de la maladie. Cette prévention passe par le dépistage de la maladie. Le diagnostic précoce permet ainsi d'agir sur certains facteurs de risque. Par exemple, équilibrer la pression artérielle ou le diabète, ou utiliser des statines ;
- prévention tertiaire : toutes les mesures visant à améliorer l'état de santé des malades en évitant les complications liées à la maladie, et en améliorant la qualité de vie des patients atteints de la maladie. Peu valorisée par les acteurs de soins, la prévention tertiaire est pourtant la médecine de demain, eu égard à l'allongement de la durée moyenne de vie, aux développements des pathologies chroniques. Par exemple, un patient coronarien doit adopter un rythme de vie, des comportements sociaux et alimentaires et/ou une médication qui diminuent le risque de nouvel infarctus. Chez le diabétique, l'hygiène de vie et le contrôle de la glycémie sont une partie importante de la prévention tertiaire.

PROPENSION (SCORE DE)

Propensity score

Le score de propension est un indicateur de la propension à utiliser une modalité thérapeutique parmi deux (parfois plus). Il est utilisé dans les études non randomisées (prospectives ou rétrospectives). Il est calculé après recherche des facteurs prédictifs du choix thérapeutique, et correspond à la probabilité de recevoir un des traitements en fonction des variables observées avant traitement. Cette probabilité est généralement issue d'une équation de régression logistique. Le score de propension cherche à évaluer avec un risque de confusion réduit l'impact des modalités thérapeutiques sur des critères comme la survie, la morbidité, les effets secondaires ou encore la qualité de vie... Il semble très pratique pour effectuer des appariements ou des stratifications afin de comparer ces critères au sein des sous-groupes obtenus. En dépit de cet artifice, qui permet d'obtenir des sous-groupes *a posteriori* similaires, cette méthode ne peut prétendre atteindre le niveau de preuve des essais randomisés, car l'absence de biais ne peut être garantie.

Le score de propension est utile en cas de faible effectif, car dans les études à large effectif, les modèles multivariés peuvent être suffisants pour produire des conclusions valides.

Si l'on considère un schéma simplifié de la situation de confusion (figure 22), trois points sont reliés : d'un côté, des facteurs dont dépendent les traitements (flèche a), mais qui sont aussi liés au résultat (flèche b). De l'autre côté, le but est de mesurer de manière non biaisée l'effet des traitements ou des interventions sur la variable étudiée.

Pour simplifier, les modèles multivariés atteignent ce but en retirant le poids de l'association entre les facteurs de confusion et le résultat tandis que le score de propension opère sur l'autre bras du triangle de confusion : il vise à annuler l'effet de l'association entre les facteurs de confusion et les traitements.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

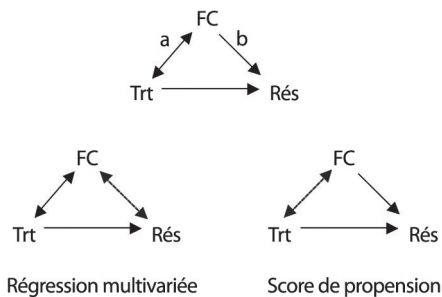


Figure 22. Schéma des types d’ajustement sur les facteurs de confusion permettant de comparer le score de propension et le modèle multivarié. FC = facteur de confusion. Trt = traitement. Rés = résultat.

PUISSANCE STATISTIQUE

➔ **BÊTA (RISQUE)**

q

QUALITÉ DE VIE

Quality of life

En plus des critères objectifs (ou bruts) comme la mortalité ou la morbidité, il est de plus en plus nécessaire d'évaluer dans les études cliniques ou même dans la pratique courante la qualité de vie (QdV) des patients. La mesure de la QdV prend en considération la perception par le patient de son propre état de santé grâce à des instruments de mesure spécifiques. Elle se fonde largement sur la définition de l'OMS (1946), qui ne définit plus la santé comme l'absence de maladie mais comme « un état complet de bien-être physique, psychologique et social ».

Cependant, la QdV appliquée à la santé, ou *Health Related Quality of Life* des Anglo-Saxons, prend en compte non pas toutes les dimensions de la QdV en général mais celles qui peuvent être modifiées par la maladie ou son traitement.

Améliorer la QdV d'un patient consiste à réduire les répercussions fonctionnelles négatives (sur l'activité physique, l'état psychologique, les relations sociales...) de sa maladie et de ses traitements, telles qu'il les perçoit. Ainsi ont été développés des instruments de mesure de la QdV dans différentes situations cliniques comme la cancérologie ou les maladies chroniques.

Le développement de questionnaires de QdV a pour objectif de rendre une entité subjective mesurable sur une échelle de valeurs numériques. On obtient ainsi un score ou un index de QdV. Le développement d'un instrument de QdV obéit à des règles établies : définition de la population cible, questions bien formulées, conditions d'administration du questionnaire, propriétés psychométriques documentées, disponibilité internationale et publication validant le questionnaire, conditions d'utilisation officielles (copyright ou domaine public), manuel d'utilisation et d'interprétation clair.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

On distingue classiquement deux types d'instruments : les instruments génériques (tels que le *Medical Outcomes Study (MOS) 36-Item Short Form*, ou *SF-36*), et les instruments spécifiques d'une pathologie donnée (par exemple, dans les maladies chroniques inflammatoires de l'intestin, l'*Inflammatory Bowel Disease Questionnaire*, ou *IBDQ*) ou d'un groupe de pathologies (tels que le *GastroIntestinal Quality of Life Index*, ou *GIQLI*, pour les maladies digestives, ou le *Quality of Life Questionnaire Core 30 Items*, ou *QLQ-C30*, et ses modules par organe en cancérologie. La liste des 562 questionnaires disponibles dans différentes spécialités médicales (mars 2007) peut être consultée sur le site du *MAPI Research Institute* (www.qolid.org).

Il est essentiel d'utiliser dans chaque pays un instrument validé (de manière psychométrique) dans la langue utilisée. Le site du *MAPI Research Institute* comporte 558 traductions validées. Certains questionnaires (mais seulement une minorité) ont été développés en langue française (par exemple, le *Reflux Qual* dans le reflux gastro-œsophagien) ou ont eu une traduction validée (comme le *SF-36* ou le *GIQLI*).

➔ **ÉCHELLES**

QUARTILES

➔ **BOÎTE À MOUSTACHES**

QUASI RANDOMISÉE (ÉTUDE)

➔ **RANDOMISATION**

QUOROM (QUALITY OF REPORTING OF META-ANALYSES)

QUOROM

Le QUOROM (*Lancet* 1999, 354 : 1896-900) est une liste d'items à laquelle doit satisfaire une méta-analyse d'essais randomisés. La liste de 17 items (tableau XV) a été élaborée au cours d'une conférence de consensus ayant réuni 30 épidémiologistes en 1996. Le QUOROM permet de juger de la qualité d'une méta-analyse d'essais randomisés et donc d'évaluer sa validité interne et externe. L'élaboration du QUOROM obéit aux mêmes règles que celles du « *CONSORT statement* » pour les essais randomisés. La liste d'items est associée à un diagramme (figure 23) permettant d'analyser le nombre d'essais retrouvés, le taux et les raisons d'exclusion d'essais.

Cette approche permet d'avoir un instrument unique pour évaluer la qualité des méta-analyses.

Tableau XV. Liste QUOROM.

Section du rapport	Descripteur	Reporté Oui/ Non	Page où l'item est rapporté
Titre	Identifier l'étude comme étant une méta-analyse ou une revue systématique d'essais (études) randomisés		
Résumé : Utiliser un style structuré			
• Buts	Détailler la question clinique posée		
• Sources de données	Donner une liste complète des bases de données consultées		
• Méthode d'analyse	Détailler les critères de sélection (par exemple, population, intervention, résultats, type d'étude), méthodes d'analyse de la validité, extraction des données, caractéristiques des études et synthèses des données quantitatives avec suffisamment de détails pour permettre une analyse similaire par d'autres chercheurs		
• Résultats	Caractéristiques des études incluses et exclues, résultats quantitatifs et qualitatifs (par exemple, estimation de l'effet, intervalles de confiance), et analyses de sous-groupes		
• Conclusions	Principaux résultats		
Introduction	Poser explicitement le problème, le rationnel physiologique ou biologique de l'intervention ainsi que celui de la revue systématique		
Méthode :			
• Recherche	Information détaillée sur les sources de données (bases de données, registres, communications personnelles, agences, recherche manuelle) et les limites de la recherche (durée, type de publication, langue)		

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Section du rapport	Descripteur	Reporté Oui/ Non	Page où l'item est rapporté
• Sélection	Fournir les critères de sélection et d'exclusion (type de population, intervention, principaux résultats, type d'étude)		
• Validité	Critères utilisés (par exemple, analyse en insu, évaluation de la qualité des études)		
• Extraction des données	Le procédé utilisé (complètement indépendant, en deux fois)		
• Caractéristiques des études	Type d'études incluses, population étudiée, détails sur l'intervention, définition des critères de jugement, analyse de l'hétérogénéité entre les études		
• Synthèse quantitative des résultats	Type de mesure de la taille de l'effet (risque relatif ou rapport des cotes), méthode de combinaison des résultats (tests statistiques et intervalles de confiance), interprétation des données manquantes, évaluation de l'hétérogénéité statistique, rationnel de l'analyse a priori de la sensibilité et des sous-groupes, évaluation d'un éventuel biais de publication		
Résultats :			
• Flux des études	Fournir un profil de la méta-analyse avec les nombres d'inclusions et d'exclusions à chaque étape		
• Caractéristiques des études	Description des données pour chaque étude incluse (par exemple, âge des patients, effectif de l'étude, intervention évaluée, dose, durée, durée de suivi)		

Section du rapport	Descripteur	Reporté Oui/ Non	Page où l'item est rapporté
• Synthèse quantitative des données	Reporter l'accord dans la sélection des études, analyse de la validité, présenter un résumé simplifié des résultats (pour chaque étude et chaque critère de jugement), présenter les données ayant permis de calculer la taille de l'effet et les intervalles de confiance sur la base d'intention de traiter (tableau 2 x 2, moyennes et écarts types, proportions)		
Discussion	Résumer les résultats clés, discuter la validité interne et externe de la méta-analyse en termes cliniques, interpréter les résultats à la lumière des données probantes (factuelles), discuter les éventuels biais de la méta-analyse (par exemple, biais de publication), et suggérer le sujet de nouvelles études		

(D'après Moher D. Cook J. D. Eastwood S. Olkim I. Rennie D. Stroup D. Improving the quality of reports of meta-analyses of randomised controlled trials : the QUORUM statement, *Lancet* 1999, 354 ; 186-1900. Avec la permission d'Elsevier Masson SAS – traduction française de l'auteur K. Slim)

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

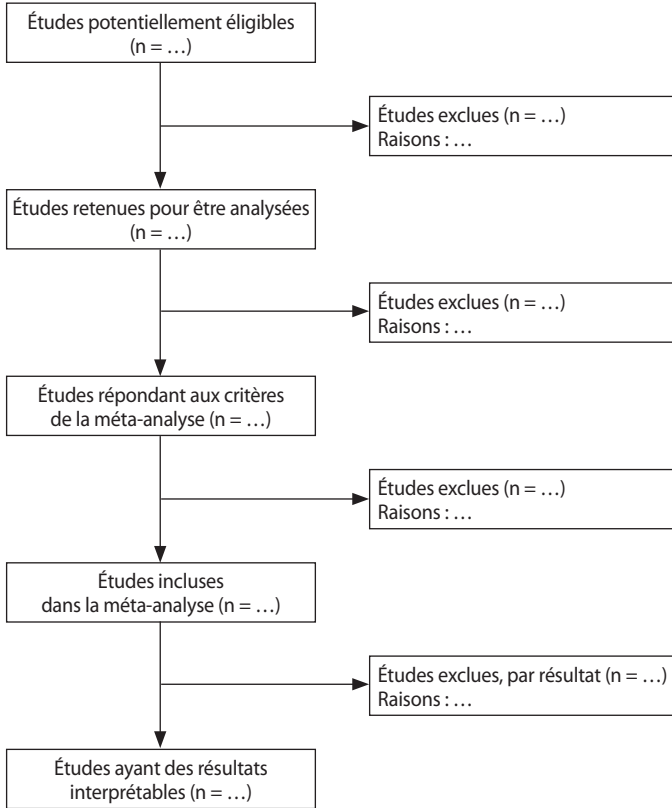


Figure 23. Diagramme des inclusions des essais (études) au fil de la méta-analyse.

(D'après Moher D. Cook J. D. Eastwood S. Olkim I. Rennie D. Stroup D. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement, *Lancet* 1999, 354 ; 186-1900. Avec la permission d'Elsevier Masson SAS – traduction française de l'auteur K. Slim)

r

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

RANDOMISATION

Randomisation

C'est la répartition des participants à une étude clinique (ou expérimentale) au hasard dans deux (ou plusieurs) groupes. La répartition ou l'attribution (allocation) est faite selon un tirage au sort aléatoire, on parle alors de randomisation systématique, ou bien selon une autre méthode plus critiquable sur le plan méthodologique (réalisée de manière alternative, selon le jour de la semaine, selon l'année de naissance, selon le praticien), et on parle alors de quasi-randomisation.

La randomisation peut se faire par personne participante (randomisation individuelle) ou par groupe d'individus (randomisation par grappe ou *cluster randomisation*).

La randomisation est la meilleure méthode pour limiter les biais statistiques car l'allocation au hasard aboutit à une répartition « idéalement » équilibrée (du fait de la loi du hasard) des différents critères, à l'exception du critère qui fait l'objet de l'étude. Quand la répartition ne peut être équilibrée, on procède à une stratification.

Exemple

Dans une étude randomisée évaluant deux méthodes thérapeutiques, les groupes de patients seront théoriquement (si l'essai est bien conçu) similaires, sauf pour la méthode thérapeutique. Donc les différences dans les résultats seront attribuées aux méthodes thérapeutiques concernées.

➔ STRATIFICATION

RANDOMISATION PAR GRAPPE

Cluster randomisation

➔ RANDOMISATION

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

RAPPORT DE HASARDS

Hazards ratio

Le rapport de hasards est le risque relatif de survenue d'un événement dans une analyse multivariée réalisée selon le modèle de régression de Cox.

Exemple

Dans une analyse de la mortalité après chirurgie colorectale, l'âge supérieur à 70 ans est associé à un risque de mortalité postopératoire de 3,9. C'est-à-dire que le patient âgé de plus de 70 ans a un risque 4 fois plus élevé de décéder après une chirurgie colique réglée.

RAPPORT DE VRAISEMBLANCE

Likelihood ratio

Dans les études diagnostiques ou épidémiologiques, le rapport de vraisemblance (RV) signifie la proportion dans laquelle la présence ou l'absence de maladie est plausible quand un test diagnostique ou un signe est présent (ou absent). Contrairement aux valeurs prédictives, le RV ne dépend pas de la prévalence d'une maladie :

- le RV positif (ou force probatoire) permet d'estimer dans quelle mesure la présence d'une maladie est possible devant un test ou un signe positif. Le RV témoigne de la relation entre la probabilité d'un test ou d'un signe positif chez les individus atteints et les individus sains : $RV + = \text{sensibilité} / (1 - \text{spécificité})$. Le RV + est automatiquement plus grand que l'unité : un test ou signe positif affirme d'autant plus le diagnostic que le RV + est proche de l'infini ;
- le RV négatif permet d'estimer dans quelle mesure la présence d'une maladie est peu possible devant un test ou un signe négatif : $RV - = (1 - \text{sensibilité}) / \text{spécificité}$. Le RV - est automatiquement plus petit que l'unité : un test ou signe négatif infirme d'autant plus le diagnostic que le RV - est proche de zéro.

Exemple

Prenons le cas d'une patiente explorée pour une anémie. En considérant son âge, l'ancienneté des symptômes, les signes cliniques et le résultat initial de la numération-formule sanguine, on estime que la probabilité d'une anémie ferriprive est de 50 %. Le test de référence pour affirmer la présence ou l'absence de cette carence est un myélogramme avec coloration spéciale. Mais il s'agit d'un test invasif. Une recherche de la littérature actuelle permet de retrouver un article de bonne qualité méthodologique : Guyatt G.H. *et al.* Laboratory diagnosis of iron-deficiency anaemia : an overview. *J Gen Intern Med* 1992, 7 : 145-53. Le dosage de la ferritine sérique de notre patiente montre une valeur de 60 mmol/L. Grâce à l'étude de Guyatt G. H. *et al.* rapportant les valeurs de sensibilité et de spécificité du test, nous pouvons calculer que le $RV + = 90\% / 15\% = 6$, et le $RV - = 10\% / 85\% = 0,12$.

Nous sommes donc obligés pour avancer dans l'étape diagnostique de réaliser un myélogramme.

RECOMMANDATIONS DE (POUR LA) PRATIQUE CLINIQUE

Guidelines for clinical practice

Les recommandations de pratique clinique (RPC) sont destinées à uniformiser les pratiques cliniques d'ordre diagnostique ou thérapeutique et sont généralement élaborées par les sociétés savantes ou les organismes de tutelle : par exemple, la Haute Autorité de Santé (HAS) en France, le *National Institute for Clinical Excellence* (NICE) au Royaume-Uni, ou le *US Guideline Clearinghouse* aux États-Unis. Elles peuvent être fondées sur les données de la littérature (elles sont alors appelées RPC factuelles) ou sur des avis d'experts ou un consensus professionnel. Elles doivent répondre à une méthodologie rigoureuse (figure 24). Elles sont souvent associées à des grades de recommandations et diffusées au sein de la communauté professionnelle. La qualité des RPC peut être évaluée par une grille méthodologique appelée *AGREE* (*Guidelines for Research and Evaluation instrument*) et dont la version française est disponible sur le site : www.has-sante.fr/portail/upload/docs/application/pdf/grille.pdf.

➔ GRADES DE RECOMMANDATION

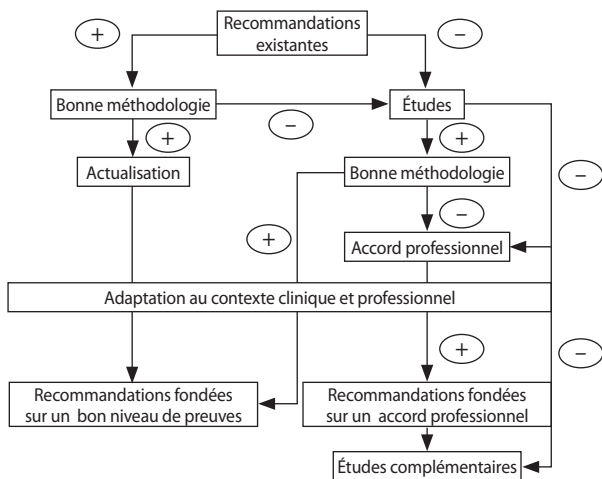


Figure 24. Méthode d'élaboration de recommandations de pratiques professionnelles selon la HAS.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

RÉDUCTION DE RISQUE

Risk reduction

La réduction de risque est qualifiée d'absolue ou relative :

- la réduction absolue de risque (RAR) est la différence entre le risque de développer une maladie (ou une complication) dans le groupe exposé (ou traité) et celui dans le groupe contrôle. Dans l'exemple fourni dans le paragraphe du rapport des cotes ou *odds ratio*, la RAR est calculée comme suit : $(a/a + b) - (c/c + d)$;
- la réduction relative de risque (RRR) est le rapport entre la différence absolue de risque entre les deux groupes et le risque dans le groupe contrôle : RAR/Rc . La RRR est le complément du risque relatif $(1 - RR)$.

➔ COTES (RAPPORT DES)

RÉGRESSION LINÉAIRE (ANALYSE DE)

Linear regression

Dans toute analyse de régression, on utilise l'association entre deux variables quantitatives afin de prédire la valeur de l'une par rapport à l'autre. En cas de régression linéaire, la relation entre deux variables continues est représentée par une ligne droite et sous la forme d'une équation simple : $y = ax + b$ (où y représente par exemple le poids à la naissance d'un enfant et x la taille de la mère). L'analyse de régression linéaire entre dans le cadre des analyses dites univariées.

RÉGRESSION LINÉAIRE MULTIPLE (ANALYSE DE)

Multiple linear regression

Dans ce cas, c'est l'association de plusieurs variables continues qui est analysée. Les variables sont classées par ordre décroissant selon l'importance de leur association avec le critère évalué. Néanmoins, plus le nombre de variables considérées est important, plus le risque de trouver une association « fortuite » est élevé. Le résultat de l'analyse devient ainsi peu pertinent.

L'analyse de régression multiple est une forme d'analyse multivariée où l'effet de facteurs dits confondants peut être corrigé.

RÉGRESSION LOGISTIQUE (ANALYSE DE)

Logistic regression

Il s'agit d'une analyse multivariée où la principale variable à expliquer est binaire (ou dichotomique).

Exemple

L'étude de l'Association française de chirurgie sur les facteurs de mortalité après chirurgie colorectale a mis en évidence (grâce à une analyse de régression multiple) une relation significative entre la mortalité postopératoire et l'âge avancé du patient (> 75 ans), la chirurgie en urgence, l'amaigrissement de plus de 10 %, et les antécédents d'accidents vasculaires cérébraux.

REVMAN®

RevMan®

RevMan®, diminutif de *Review Manager*, est un logiciel développé par la Cochrane Collaboration pour aider les chercheurs à réaliser une revue systématique ou une méta-analyse selon une méthode standardisée. Toutes les revues Cochrane publiées ont été menées grâce à ce logiciel. Il peut être téléchargé gratuitement sur le site : www.cc-ims.net/RevMan/download.htm.

REVUE – REVUE SYSTÉMATIQUE

→ SYNTHÈSE, SYNTHÈSE MÉTHODIQUE

RISQUE ABSOLU

→ NOMBRE DE SUJETS À TRAITER

RISQUE RELATIF

Relative risk

Le risque relatif (RR) est le rapport entre le risque de survenue d'un événement (maladie, décès, guérison, complication) dans le groupe traité ou évalué et celui dans un groupe contrôle. Dans une étude épidémiologique de cohorte, le RR permet d'estimer la probabilité de survenue d'une maladie dans le groupe exposé à un facteur de risque, le risque y étant autant de fois supérieur ($RR > 1$) ou inférieur ($RR < 1$) à celui du groupe témoin. Dans un essai thérapeutique, le RR permet d'estimer la probabilité que le résultat dans le groupe traité soit autant de fois supérieur ou inférieur à celui dans le groupe contrôle.

Par convention, un $RR < 1$ témoigne d'un effet bénéfique tandis qu'un $RR > 1$ témoigne d'un effet délétère. Un $RR = 1$ suggère l'absence d'effet (qu'il soit bénéfique ou délétère) d'un traitement. Par exemple, un RR de 0,5 signifie que le risque est divisé par 2 ; quand le $RR = 3$, le risque est multiplié par 3.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Le calcul du risque relatif est fait à partir d'un tableau de contingence 2×2 et est égal au rapport entre la fréquence de l'événement dans le groupe étudié et la fréquence dans le groupe témoin, soit $a/(a + b)/c/(c + d)$.

Le RR est une valeur sans unité et ne renseigne pas directement sur l'importance de l'événement et la pertinence clinique du résultat.

Exemple 1

Si on considère que R0 est le risque dans une population témoin et R1 le risque dans la population étudiée, l'importance du RR dépend en fait de la fréquence de l'événement :

- pour R0 = 2/10 000 avec R1 = 1/10 000, le RR = 0,5, soit 50 % d'amélioration d'un événement très rare ;
- pour R0 = 200/10 000 avec R1 = 100/10 000, le RR = 0,5, soit 50 % d'amélioration d'un événement beaucoup plus fréquent.

Exemple 2

Deux groupes de 8 patients sont comparés. Dans le groupe A, 2 patients ont présenté l'événement (prenons l'exemple d'une complication chirurgicale), d'où une fréquence de l'événement de 2/8, soit 0,25. Dans le groupe B, 5 patients ont présenté l'événement, soit une fréquence de 5/8, soit 0,62. Le risque relatif de survenue de la complication est de 0,62/0,25, soit 2,48. Donc le groupe B a 2,5 fois plus de présenter l'événement (en l'occurrence, une complication chirurgicale), le traitement du groupe B doit être abandonné au profit du groupe A.

Pour conclure, le risque relatif, le rapport des cotes (*odds ratio*), les réductions de risque, et le nombre de sujets à traiter (NST) ne véhiculent pas la même information clinique. Ainsi, ces mesures ne donnent pas exactement les mêmes renseignements sur la pertinence clinique d'un effet. Les deux premières (rapport des cotes, risque relatif) sont des mesures relatives et estiment un bénéfice relatif, tandis que les deux dernières (différence de risque et NST) mesurent un bénéfice absolu :

- le bénéfice relatif est plutôt une information explicative. Il est le reflet direct de l'efficacité du traitement. C'est une information qui intéresse le chercheur ;
- le bénéfice absolu reflète plus les conséquences apportées par un traitement au niveau d'une population. Il est spécifique d'une situation particulière : traitement caractérisé par son risque relatif, type de patients conditionnant le niveau de risque de base, durée de traitement ou de suivi. Il est plus pertinent en termes de santé publique.

Une réduction relative de risque de 30 % est déjà une réduction conséquente. Mais la pertinence clinique de cet effet dépend du risque de

base. Si le risque de base est de 50 %, sous l'effet d'une réduction de 30 %, il devient de 35 %, donnant une différence de risque de 15 %. Avec un risque initial de 5 %, la même réduction relative aboutit à un risque sous traitement de 3,5 %, correspondant à une différence absolue de 1,5 %. En termes d'événements évités pour 1 000 sujets traités (NST), le premier cas de figure correspond à 150 événements évités, tandis que le second à seulement 15. Aussi bien du point de vue de la santé publique que du point de vue individuel, la première situation est plus intéressante que la seconde.

➔ **COTES (RAPPORT DES)**

ROC

➔ **COURBE ROC**

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

S

SCOPUS

SCOPUS est une base de données électronique ayant succédé à EMBASE, éditée par Elsevier, et accessible sur le site www.scopus.com. SCOPUS couvre le sommaire d'environ 15 000 revues par plus de 4 000 éditeurs de 1974 à nos jours.

Cette base de données contient près de 33 millions de résumés et 386 millions de pages web scientifiques. Elle permet aussi de suivre les citations des articles référencés. L'accès à SCOPUS n'est pas gratuit mais nécessite un abonnement.

SCIENCE CITATION INDEX

Science Citation Index

Le *Science Citation Index* est une base de données créée par l'institut américain ISI (*Institute for Scientific Information*, www.scientific.thomson.com/isi/). Cet institut privé tient à jour un index de citations des articles scientifiques publiés dans le domaine des sciences sociales et humaines. En bibliométrie, cet index permet de juger de l'« importance scientifique » d'un article mais aussi d'un auteur ou d'une revue. L'ISI publie chaque année le *Journal Citation Reports* (JCR), qui évalue la majorité des revues scientifiques (près de 5 700 revues en sciences et technologies dans l'édition JCR Web Science). On peut ainsi connaître les revues les plus citées dans un domaine particulier, et celles qui ont le meilleur facteur d'impact (ou *Impact Factor*, proposé il y a près de 50 ans par Eugène Garfield comme un indicateur de la « consommation scientifique ». Il s'agit de calculer pour une année donnée le rapport entre le nombre de citations en numérateur et le nombre d'articles publiés en dénominateur pendant les deux années précédentes). Les autres indices fournis dans le JCR sont l'indice d'immédiateté (*immediacy index*), la demi-vie des

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

a citations (*cited half-life*), le rang de classement de la revue (*journal ranking*), l'indice d'autocitations (*self citing index*), le nombre total de citations, et le nombre de pages électroniques consultées (*hit rates*). Une discussion de l'importance et des limites du facteur d'impact et des autres index a été récemment publiée (Mariette C., Slim K. *J Chir* 2006, 143 : 275).

e **SECRET D'ATTRIBUTION**

f *Concealment of allocation*

g Tout essai randomisé thérapeutique (mais aussi diagnostique) comporte une comparaison entre un (des) groupe(s) traité(s) et un groupe contrôle. Le tirage au sort aboutit à l'attribution aléatoire du traitement étudié ou du traitement contrôle. Pour préserver l'objectivité de l'analyse, la constitution des groupes doit être faite en aveugle, c'est-à-dire qu'au moment où le patient est inclus dans le tirage au sort, personne ne doit être au courant *a priori* du type de traitement qui lui sera alloué. La méthode respectant le secret d'attribution consiste à préparer des enveloppes cachetées ou des codes avant le début de l'essai. Le groupe dans lequel sera inclus le patient n'est connu qu'au dernier moment après tirage au sort. Cela évite de connaître le type de traitement qui sera alloué et d'être tenté de ne pas inclure ou de décaler le patient à sa convenance, ce qui aboutit obligatoirement à un « biais d'attribution ». Le secret d'attribution est un critère essentiel de qualité d'un essai randomisé. Les essais qui ne respectent pas cette obligation sont appelés essais « quasi randomisés ».

q **Exemple de non-respect du secret d'attribution**

r Si le chercheur sait le type de traitement qui sera alloué au prochain patient à inclure (allocation alternée, allocation en fonction de la date d'inclusion – jours pairs-jours impairs – ou de l'année de naissance du patient, enveloppes non cachetées), il peut être tenté de ne pas l'inclure ou de choisir l'enveloppe à sa convenance.

s **➔ BIAIS**

t **SENSIBILITÉ**

u *Sensitivity*

v La sensibilité est un indice informationnel sur la probabilité d'avoir un test diagnostique positif ou un signe clinique chez un patient ayant la maladie. La sensibilité (rapport des vrais positifs sur l'ensemble des sujets malades) reflète le taux de vrais positifs et est inversement proportionnelle au taux de faux négatifs. Une sensi-

bilité de 100 % signifie que le test diagnostique est constamment positif ou que le signe est constamment présent (il n'y a aucun faux négatif), ce qui reste exceptionnel en médecine.

SÉRIE DE CAS

Case series

C'est une étude d'observation incluant d'une manière consécutive ou non un groupe de patients ayant une maladie ou ayant subi un traitement particulier. Elle peut être prospective ou rétrospective. Elle permet généralement d'émettre une hypothèse qui devra être confirmée (ou non) par une étude contrôlée (randomisée ou non). La littérature foisonne de ces études sous la forme de petites séries de cas, dont les conclusions doivent être interprétées avec précaution.

SIMPLE AVEUGLE

➔ AVEUGLE

SOMMATION

Pooling

Il s'agit de l'intégration, dans une même analyse, des résultats de plusieurs études de cohorte ou essais contrôlés. Cette sommation peut être utilisée dans les revues systématiques sous la forme d'une simple addition de cas ou d'événements sans pondération ni évaluation de la taille de l'effet comme dans les vraies méta-analyses. La sommation donne une indication qui devra faire l'objet d'une méta-analyse méthodologiquement rigoureuse si les données de la littérature le permettent.

Exemple

L'évaluation de la morbidité sur un grand nombre de patients peut être faite sous la forme d'une sommation des événements rapportés dans la littérature, comme dans la revue systématique de Yong L. *et al.* (*Surg Endosc* 2001, 15 : 1431-9) sur les colectomies par coelioscopie pour cancer, à comparer avec la vraie méta-analyse de Reza M. M. *et al.* (*Br J Surg* 2006, 93 : 921-8) sur le même sujet.

SOURCES PRIMAIRES

Primary sources

Dans le schéma de la médecine factuelle, la recherche des données de la littérature constitue la deuxième étape après celle de la question posée (d'ordre diagnostique, thérapeutique ou pronostique)

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

a (voir tableau VIII). La recherche peut être faite dans les sources dites
b « primaires », qui fournissent la liste des études à lire pour obtenir
c la réponse à la question posée (approche appelée *doing mode*). La
d principale source de revues systématiques et d'essais randomisés est
e la librairie Cochrane. Une manière aisée de consulter les sources
f factuelles primaires est de faire une recherche préliminaire dans la
g base de données TRIP (www.tripdatabase.com).

h Les autres sources sont : Medline, Scopus, Cinahl (base de données
i destinée aux professions paramédicales : www.cinahl.com/library/library.htm), PsycLIT (destinée aux spécialités relatives à la psy-
j chiatry ou la psychologie : www.apa.org/psycinfo/psycinfo.html),
k CancerLIT (pour la cancérologie, incluse dans www.pubmed.org),
l ou Google Scholar (www.scholar.google.com).

m Les limites de ces sources sont essentiellement la complexité de la
n recherche et surtout le fait qu'il faut sélectionner et lire les articles
o retrouvés, interpréter leurs résultats, et évaluer leur validité interne
p et externe. Ces étapes nécessitent une expertise méthodologique cer-
q taine, d'où le recours aux sources dites secondaires.

r ➔ **COCHRANE, MEDLINE, SCOPUS**

s **SOURCES SECONDAIRES**

t *Secondary sources*

u L'avantage des sources secondaires est qu'elles procurent une infor-
v mation préalablement validée par des experts selon le concept de
w la médecine factuelle (approche appelée *using mode*). Internet com-
x porte une multitude de sources dont une liste non exhaustive peut
y être consultée sur le site de la SchARR (www.shef.ac.uk/scharr). Mais
z les internautes doivent utiliser ces sources avec un esprit critique
pour pouvoir séparer le bon grain de l'ivraie. Pour être considérée
comme fiable, toute source secondaire doit répondre à certains cri-
tères. Elle doit :

- faire état de l'absence de conflit d'intérêt ;
- exprimer clairement le sujet traité ;
- adopter une méthodologie de recherche bibliographique rigou-
reuse et clairement formulée ;
- tenir une mise à jour régulière.

Le tableau XVI résume les différentes sources secondaires en fonc-
tion du type de données qu'elles peuvent fournir.

Tableau XVI. Exemples de sources secondaires de la médecine factuelle, classées selon le type de données fournies.

Type	Caractéristiques	Nom	Adresse Internet
Données par question clinique (<i>Critically Appraisal Topics</i> ou CAT)	Fournit des réponses à des questions cliniques précises	CAT Crawler	www.bestbets.org/index.html
		Club de Journaux	www.acpjc.org
Résumés factuels	Revue des données factuelles sur une question clinique donnée	Bandolier	www.jr2.ox.ac.uk/bandolier
		Clinical Evidence	www.clinicalevidence.com
Résumés structurés et commentés	Analyse des plus importants articles publiés	EBM Online	www.ebm.bmj.com
		ACP Journal Club	www.acpjc.org
		EBM-Journal	http://phenix.itnetwork.fr
Évaluations technologiques	Évaluation de la qualité des preuves scientifiques concernant une intervention donnée	NHS-HTA (Royaume-Uni)	www.hta.nhsweb.nhs.uk
		HAS (France)	www.has-sante.fr

SOUS-GROUPES (ANALYSE DE OU EN)

Subgroups (analysis)

Les analyses de sous-groupes (groupes particuliers de patients traités ou exposés à un risque) viennent souvent compléter le résultat principal d'un essai clinique. Il s'agit de subdiviser de manière prospective et avant le début de l'essai la population d'un essai en deux ou plusieurs sous-groupes et à étudier l'efficacité du traitement dans chacun de ces sous-groupes.

Exemple

Une analyse en sous-groupe suivant le sexe revient à mesurer séparément l'effet du traitement chez les hommes et chez les femmes. Une analyse suivant le stade d'une maladie revient à évaluer l'efficacité d'un traitement en fonction de la gravité de cette maladie, par exemple, l'efficacité d'une radiothérapie préopératoire en fonction du stade T d'une tumeur du rectum (Kapiteijn E. *et al.* Preoperative radiotherapy combined with total mesorectal excision for respectable rectal cancer. *N Engl J Med* 2001, 345 : 638-46).

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Dans un essai qui n'a pas montré de différence statistiquement significative (essai non concluant ou essai négatif), le but des analyses en sous-groupes est de rechercher le ou les sous-groupes dans lesquels existerait un effet du traitement statistiquement significatif.

Dans un essai concluant (ou positif, où une différence statistiquement significative a été obtenue), le but serait alors de rechercher le ou les sous-groupes dans lesquels le traitement serait le plus efficace et surtout ceux dans lesquels il serait inefficace. L'objectif est de mieux définir la population cible en restreignant éventuellement la diffusion du traitement par rapport à la population qui a été incluse dans l'essai.

Ainsi, dans l'exemple cité plus haut, l'essai de Kapiteijn E. *et al.* a montré que la radiothérapie avant résection totale du mésorectum n'est pas utile pour les tumeurs au stade précoce.

Une bonne illustration des analyses de sous-groupes est l'essai stratifié. Ces essais étudient simultanément deux ou plusieurs strates, qui sont l'équivalent des sous-groupes. La méthodologie de ces essais comporte :

- la formulation explicite des hypothèses concernant les différentes strates dans l'objectif de l'essai ;
- le calcul du nombre de sujets nécessaires pour chacune des strates ;
- la randomisation indépendante par strate ;
- la prise en compte de la multiplicité des tests statistiques.

Exemples

- Dans un essai sur la place de la coelioscopie dans les résections coliques pour cancer, les résultats peuvent être analysés séparément dans le sous-groupe colectomies droites et dans le sous-groupe colectomies gauches.
- Dans un essai sur la prévention des fistules anastomotiques après résection pancréatique, une analyse par sous-groupes comportera le sous-groupe de résection sur parenchyme pancréatique mou et le sous-groupe de résection sur parenchyme scléreux. La randomisation est alors réalisée séparément pour chaque strate ou sous-groupe.

Dans les méta-analyses, la constatation d'une hétérogénéité, par exemple, implique (après élimination du ou des essais induisant l'hétérogénéité) une analyse des sous-groupes à la recherche d'interactions.

Mais l'analyse des sous-groupes peut être faite en dehors de toute hétérogénéité. Ces sous-groupes sont définis en fonction :

- du type de traitement ;

- du type de mesure du critère de jugement (par exemple la douleur, avec l'échelle visuelle analogique, la consommation d'antalgiques) ;
- des caractéristiques des patients ;
- des caractéristiques des essais.

Mais même définie *a priori*, l'analyse de sous-groupes, dans les méta-analyses, est obligatoirement faite *a posteriori* (analyse *post-hoc*). Son but est alors d'évaluer la sensibilité (selon le même principe que pour les essais thérapeutiques), de générer une hypothèse et de la vérifier. Les risques de la multiplication des analyses de sous-groupes dans les méta-analyses (comme dans les essais randomisés), sont essentiellement liés à la possibilité d'une erreur de première espèce ou erreur α . La multiplication des tests statistiques augmente le risque d'obtenir un test positif uniquement du fait du hasard.

SPÉCIFICITÉ

Specificity

La spécificité est un indice informationnel sur la probabilité d'avoir un test diagnostique négatif ou l'absence d'un signe clinique chez un sujet indemne. La spécificité (rapport des vrais négatifs sur l'ensemble des sujets indemnes) reflète le taux de vrais négatifs et est inversement proportionnelle au taux de faux positifs. Une spécificité de 100 % permet d'affirmer qu'un signe est pathognomonique (il n'y a aucun faux positif). Par exemple, des vomissements fécaloïdes dans une occlusion intestinale ont une spécificité proche de 100 %, et affirment l'occlusion (les fistules gastrocoliques étant actuellement très exceptionnelles).

STANDARDISATION

Standardization

La standardisation (ou ajustement) est une méthode épidémiologique pour comparer deux taux. En épidémiologie, comparer deux taux bruts sans tenir compte des caractéristiques individuelles des groupes peut être inadéquat si on ne tient pas compte des facteurs de confusion qui diffèrent entre les groupes (l'âge, le sexe, le niveau socioéconomique). Dans la pratique, l'âge est un facteur de standardisation ou d'ajustement classique, mais d'autres critères peuvent être utilisés pour définir des sous-groupes. Deux méthodes sont utilisées, la standardisation directe et la standardisation indirecte :

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

- dans la méthode de standardisation directe, on calcule les taux que l'on s'attendrait à trouver dans les populations étudiées si elles avaient toutes la même composition selon la variable que nous voulons ajuster ou contrôler (comme, par exemple, l'âge, le groupe socioéconomique, ou d'autres caractéristiques). On utilise la structure d'une population type, appelée « standard », stratifiée selon la variable que l'on veut contrôler, et à laquelle on applique les taux spécifiques des strates correspondantes dans la population étudiée. On obtient ainsi le nombre de cas « attendus » dans chaque strate si les populations avaient la même composition ;
- la standardisation indirecte diffère dans la méthode et dans son interprétation. Au lieu d'utiliser la structure de la population type, on utilise ses taux spécifiques et on les applique aux populations à comparer, préalablement stratifiées selon la variable à contrôler. De cette façon, on obtient le total de cas attendus.

Le taux de morbidité ou de mortalité standardisé ou ajusté (*Standardized Morbidity, Standardized Mortality Ratio* ou *SMR*) est le rapport des morbidités ou des mortalités standardisées. Il est défini comme le rapport du nombre de cas observés sur le nombre de cas attendus.

Il arrive souvent que le SMR soit exprimé en pourcentage ; il faut alors multiplier le SMR précédemment donné par 100 pour l'obtenir. Par exemple, un SMR de 2 ou de 200 % signifie que la mortalité dans le groupe étudié est le double de celle de la population type.

Le terme SIR (*Standardized Incidence Ratio*) est souvent employé lorsqu'on s'intéresse à l'incidence et non à la mortalité.

STARD (STANDARDS FOR REPORTING OF DIAGNOSTIC ACCURACY)

STARD

Comme le CONSORT pour les essais randomisés ou le QUOROM pour les méta-analyses, le STARD a été développé (www.consort-statement.org/Initiatives/newstard.htm) pour améliorer la qualité des publications sur les moyens diagnostiques. Le but est de permettre au lecteur d'évaluer le risque de biais dans une étude, la reproductibilité et la « généralisabilité » des résultats obtenus. Les annexes 1 et 2 présentent les différents items du STARD. Ce standard a été adopté par plusieurs journaux : *Clin Chemistry, Ann Intern Med, Radiology, Br Med J, Lancet, Am J Clin Pathol, Clin Biochemistry, Clin Chem Lab Med.*

STATISTIQUEMENT SIGNIFICATIF

Statistically significant

Ce terme signifie que la différence observée est suffisamment importante par rapport aux fluctuations aléatoires, c'est-à-dire que le résultat n'est pas attribué au hasard ; dans un processus de validation, un tel résultat est interprété comme invalidant l'hypothèse nulle (l'hypothèse nulle signifie que le résultat est dû au hasard) vérifiée par des tests statistiques adaptés. Le terme statistiquement significatif est généralement ou « consensuellement » associé à une différence pour laquelle la valeur p est inférieure à 0,05. Cela veut dire qu'il y a moins de 5 chances sur 100 que la différence soit due au hasard. Le seuil de significativité est donc un moyen de contrôler le risque d'erreur alpha. Il ne prend pas directement en compte le risque d'erreur bêta. Il faut noter que le seuil de significativité statistique à 0,05 peut être abaissé (par exemple à 0,01) selon le contexte de l'étude.

Exemple de seuil à 0,05 (ou 5 %) peu pertinent

Voir à « Hypothèse nulle », l'exemple des spécialités pharmaceutiques dans le Vidal®.

Par ailleurs, il faut distinguer significativité statistique et pertinence clinique car une différence statistiquement significative n'a pas obligatoirement un impact clinique pour les patients.

Exemple de différence statistiquement significative mais cliniquement peu pertinente

Une différence statistiquement significative dans l'évaluation de la douleur postopératoire après deux techniques chirurgicales évaluée à 0,5 sur l'échelle visuelle ou à 5 mg de morphine n'a pas nécessairement une signification clinique pertinente.

STRATIFICATION

Stratification

Dans un essai randomisé, les strates, ou sous-groupes (en fonction de critères propres à chaque étude), sont définis *a priori* pour augmenter la puissance des tests statistiques en diminuant la variance du critère de jugement. Cela implique des tirages au sort séparés fondés sur deux ou plusieurs tables.

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

P

q

r

s

t

u

v

w

x

y

z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Exemple

Dans un essai sur l'efficacité de la chimiothérapie dans les cancers colorectaux, on peut stratifier le tirage au sort en fonction de la présence ou non de métastases hépatiques synchrones.

➔ SOUS-GROUPES (ANALYSE DE OU EN)

STROBE (STRENGTHENING THE REPORTING OF OBSERVATIONAL STUDIES IN EPIDEMIOLOGY)

STROBE

L'initiative STROBE (www.strobe-statement.org/index.html) a été lancée, par des journaux comme *Ann Intern Med*, *Br Med J*, *Bull WHO*, *Int J Epidemiol*, *JAMA*, *Sozial- und Präventivmedizin*, *Lancet*, pour améliorer la qualité des articles sur les études « observationnelles » (non randomisées) en épidémiologie réalisées par les mêmes structures qui ont développé le CONSORT, le QUOROM, le STARD ou le TREND. Le STROBE comporte 22 items (annexe 3).

STUDENT (TEST T DE)

➔ TESTS STATISTIQUES

SURVIE

Survival

En cancérologie (ou même en cas de maladies infectieuses), plusieurs types de survie sont calculés :

- survie à 5 ans : la courbe de survie étant plate au-delà de 5 ans pour une majorité de cancers, le taux de survie à 5 ans est la donnée le plus souvent utilisée dans les études de survie. C'est le taux de patients en vie 5 ans après l'intervention thérapeutique ou l'inclusion dans une étude ;
- survie globale : c'est le taux de patients en vie au terme d'une étude quelle que soit la cause du décès (du fait de la maladie elle-même, ou de la toxicité du traitement) ;
- survie médiane : c'est la durée de suivi au terme de laquelle la moitié des patients sont décédés (ou vivants). Ce critère, qui permet d'évaluer rapidement un nouveau traitement, ne permet pas de préjuger combien de sujets seront réellement guéris grâce à ce traitement ;

- survie sans progression (*progression-free survival*) : c'est la durée de survie avec la preuve objective (sur des données morphologiques) de stabilisation du volume tumoral. La survie sans progression est généralement associée à l'absence d'altération de la qualité de vie. Dans le cadre des traitements palliatifs, il est souvent nécessaire d'utiliser des protocoles de chimiothérapie en deuxième ligne. L'évaluation de ces protocoles fait appel au rapport des survies sans progression de la 2^e ligne sur la 1^{re} ligne chez un même patient. Ce rapport est appelé le *Growth Modulation Index* (GMI). Un GMI supérieur à 1 indique que la 2^e ligne a une excellente activité ;
- survie sans récurrence ou sans maladie (*disease-free survival*) : c'est la durée de survie avant la première récurrence de la maladie (par exemple, au cours d'un traitement adjuvant). La prolongation de la survie sans maladie ou sans récurrence peut ne pas être associée à une prolongation de la survie globale, mais il s'agit d'un critère important en cancérologie car la durée de survie sans maladie ou sans récurrence est généralement associée à une meilleure qualité de vie.

SYNTHÈSE, SYNTHÈSE MÉTHODIQUE

Review, systematic review

Une synthèse est une revue de la littérature sur un sujet grâce au recueil des études publiées. La synthèse peut être classique, dite aussi narrative ou qualitative, où l'auteur procède à une sélection des études à retenir sans stratégie formelle ou préalablement définie. Il s'agit alors d'une synthèse qui ne répond pas aux normes de la médecine factuelle.

Ailleurs, c'est une synthèse méthodique, essayant de répondre à une question précise, où la recherche de données obéit à une stratégie définie et rigoureuse, et où les résultats sont rapportés et analysés de manière explicite. Le tableau XVII résume les caractéristiques d'une synthèse méthodique. La méta-analyse est toujours précédée par une synthèse méthodique.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

Tableau XVII. Critères méthodologiques de qualité d'une synthèse méthodique de la littérature.

1	La synthèse évoque un problème clinique précis
2	Description de la méthode de recherche bibliographique
3	Utilisation de critères précis pour la sélection des articles
4	Évaluation de la qualité méthodologique et de la validité des études incluses
5	Reproductibilité de l'évaluation des articles (faite par un deuxième lecteur)
6	Proposition d'orientation des futures recherches
<i>Pour les synthèses comportant des recommandations thérapeutiques</i>	
7	Discussion des causes des éventuelles hétérogénéités des résultats des études
8	Réalisation d'une synthèse quantitative des données
9	Utilisation de critères de jugement cliniquement pertinents
10	Discussion de la reproductibilité des résultats

t

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

TABLES DE NOMBRES ALÉATOIRES

Random number tables

Ce sont des tables utilisées pour répartir de manière aléatoire les sujets inclus parmi deux ou plusieurs bras, afin de produire un tirage au sort équitable. La construction de ces tables n'est pas faite par l'homme (du fait d'un risque de subjectivité ; par exemple, il a été vérifié que dans le choix d'un nombre de 1 à 10, le 7 est très souvent nommé) mais par des moyens purement mécaniques, comme le décosaédrique portant sur ses faces deux fois les chiffres de 0 à 9, la roulette à 10 secteurs ou l'ordinateur.

Exemple

Voici une liste de chiffres aléatoire de 0 à 9 générée par un ordinateur : 08 96 68 56 58 06 13 47 97 59 04 85 96 49 47 43 53 24 93 09. Il suffit de décider que les patients ayant eu un chiffre pair auront le traitement A, et ceux qui ont eu le chiffre impair auront le traitement B.

Plusieurs générateurs de tables sont disponibles sur Internet (par exemple : <http://perso.orange.fr/jpq/proba/tablealea/index.htm>).

TECHNIQUE DELPHES

Delphi technique

La technique ou méthode Delphes est une méthode visant à organiser la consultation d'experts sur un sujet précis, souvent avec un caractère prospectif important. Cette méthode tient son nom de la légende de l'oracle d'Apollon Pythien à Delphes.

Cette méthode sert principalement à dégager un consensus sur les tendances futures parmi un groupe d'experts, mais elle peut aussi appuyer la prise de décision collective au sujet d'une grande variété de questions. Cette méthode repose sur une série ordonnée de discussions planifiées entre des participants, dirigées par un animateur.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

La participation est limitée à des experts ou à des personnes ayant un intérêt commun pour une question précise, et les groupes sont organisés de façon que chaque participant apporte un éclairage unique à la démarche.

La technique Delphes sert à prendre des décisions et à élaborer des stratégies fondées sur des faits, reflétant l'opinion des experts sur des questions bien définies. Cette démarche, qui n'est pas exclusive au domaine médical, est particulièrement utile pour connaître l'opinion des experts à distance, sans organiser de rencontres en personne. Étant donné que les commentaires sont faits de façon anonyme avec la technique Delphes, les risques que les idées de certains l'emportent sur d'autres sont réduits.

Reconnue comme la technique de consensus la plus utilisée de ces 40 dernières années, l'application de la méthode Delphes demeure encore confidentielle en Europe, probablement du fait de ses contraintes et limites :

- elle est relativement lourde et fastidieuse tant pour les analystes que pour les experts (4 tours de questionnaire) ;
- elle apparaît à certains égards davantage intuitive que rationnelle ;
- seuls les experts qui sortent de la norme sont amenés à justifier leur position ;
- elle suppose une excellente capacité des analystes au niveau des traitements des réponses et de la conduite maîtrisée de tout l'exercice.

TESTS STATISTIQUES

Statistical tests

Le principe essentiel des tests statistiques repose sur le rejet de l'hypothèse nulle (absence de différence entre les deux éléments comparés) au profit d'une hypothèse alternative. Pour cela, il convient de choisir le meilleur test en fonction de la nature de la variable étudiée et de sa distribution, gaussienne (normale) ou non. On classe schématiquement les tests statistiques en deux groupes : les tests paramétriques et les tests non paramétriques.

Les tests les plus usuels sont, par ordre alphabétique :

- test de Cochran-Mantel-Haenszel : c'est un test du Chi^2 recherchant la solidité d'une association entre deux variables. C'est un test sur les rangs et il est adapté aux tables de contingences ;
- test exact de Fisher : ce test est utilisé en cas de petits échantillons ne permettant pas de faire un test du Chi^2 ;

- test de Friedman : c'est un test non paramétrique utilisé en cas de deux facteurs sans répétition ;
- test de Kruskal-Wallis : c'est une analyse non paramétrique de variance (ANOVA), dérivée du test de Mann-Whitney, mais portant sur une variable qualitative à plus de deux classes ;
- test U de Mann-Whitney : c'est un test non paramétrique permettant de calculer la différence entre deux groupes (ce test est alors l'équivalent du test t de Student). Par ailleurs, ce test est aussi similaire au test de Wilcoxon ;
- test de Mantel-Haenszel : c'est un test du Chi^2 analysant la relation entre deux variables nominales après contrôle des autres variables ; Il calcule l'importance de l'association entre deux variables ;
- test de McNemar : ce test est utilisé pour l'analyse de données appariées ;
- tests non paramétriques : contrairement aux tests paramétriques, ces tests ne donnent pas d'importance à la distribution d'une variable (gaussienne ou non) ; ce sont, par exemple, les tests de Mann-Whitney et le test de Kruskal-Wallis ;
- tests paramétriques : contrairement aux tests non paramétriques, ces tests sont utilisés uniquement en cas de distribution normale (gaussienne) ; ce sont, par exemple, les tests t de Student et l'analyse paramétrique de variance (ANOVA) ;
- test Q de Cochran : ce test est utilisé dans les méta-analyses pour étudier l'hétérogénéité ;
- test t de Student : ce test paramétrique est utilisé dans les comparaisons de moyennes de deux groupes. Il nécessite que les variables aient une distribution normale (gaussienne) et que les variances des deux groupes soient similaires. Le calcul de p (à partir d'une table statistique) permet alors de voir si la différence est statistiquement significative. Il est généralement utilisé en cas de faibles effectifs ($N \leq 30$) ;
- test de Wilcoxon : ce test non paramétrique, appelé aussi test de la somme des rangs (*Wilcoxon rank sum test*), est similaire au test U de Mann-Whitney ;
- test du χ^2 (Chi^2) : ce test est utilisé en cas de variables nominales ou catégorielles quand on veut analyser la réelle différence entre deux (ou plus) proportions observées. La comparaison est faite entre les valeurs observées et les valeurs attendues sur la base de l'hypothèse nulle d'indépendance. Le test du Chi^2 permet de calculer, à l'aide d'une table statistique, la valeur de p qui permet de voir si la différence est statistiquement significative (qui n'est pas

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

due au hasard). Les tests de Mantel-Haenszel, de Cochran-Mantel-Haenszel et du χ^2 de Pearson sont des tests du χ^2 . Les tableaux XVIII et XIX résumant schématiquement l'utilisation en pratique des tests statistiques les plus usuels.

Tableau XVIII. Tests paramétriques en fonction des variables étudiées et des effectifs.

	Variable	
	Qualitative	Quantitative
2 échantillons ou groupes de patients indépendants	χ^2	t de Student
> 2 échantillons ou groupes de patients indépendants	χ^2	Analyse de variance (ANOVA) à 1 facteur
2 échantillons ou groupes de patients appariés	χ^2 de McNemar	Test t des séries appariées ou analyse de variance à 2 facteurs (ANOVA)
> 2 échantillons ou groupes de patients appariés	χ^2 de Mantel-Haenzel	Analyse de variance des mesures répétées (ANOVA)

Tableau XIX. Tests non paramétriques en fonction des variables étudiées et des effectifs.

	Variable	
	Qualitative	Quantitative
2 échantillons ou groupes de patients indépendants	Test exact de Fisher	Test de Wilcoxon ou Mann-Whitney
> 2 échantillons ou groupes de patients indépendants	Test χ^2 corrigé Yates	Test de Kruskal-Wallis
2 échantillons ou groupes de patients appariés	χ^2 de McNemar	Test t des séries appariées ou test de Wilcoxon
> 2 échantillons ou groupes de patients appariés	Test Q de Cochran ou χ^2 de Mantel-Haenzel	Test de Friedman

TREND (TRANSPARENT REPORTING AND EVALUATIONS WITH NON-RANDOMIZED DESIGNS)

TREND

Dans le même esprit que le *CONSORT statement* pour les essais randomisés, le TREND est une liste d'items visant à améliorer la qualité de publication des études non randomisées. Il est similaire au STROBE, qui reste spécifique aux études épidémiologiques. La liste comporte aussi 22 items et est disponible sur Internet (www.trend-statement.org/asp/trend.asp).

Contrairement à d'autres instruments d'évaluation comportant parfois des scores, comme la liste Nottingham-Ottawa (www.ohri.ca/programs/clinical_epidemiology/oxford_web.ppt#9), la liste de Downs-Black (*J Epidemiol Community Health* 1998, 52 : 377-84), ou l'index MINORS (Slim K. *et al. ANZ J Surg* 2003, 73 : 72-6), la liste TREND ne permet pas de faire une évaluation chiffrée de la qualité des études non randomisées incluses dans une méta-analyse. Une deuxième version du TREND a été publiée mais elle est encore en cours de validation.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z

V

VALIDITÉ

Validity

La validité d'un test, d'un examen, ou d'un instrument signifie qu'il mesure effectivement ce qu'il est censé mesurer. La validité d'une étude indique le degré de justesse de ses conclusions.

Dans le cadre de la recherche clinique et pour évaluer une étude, on parle de validité interne, validité externe, et validité de critère :

- validité interne (*internal validity*) : elle est évoquée lorsque les résultats de l'étude peuvent être rapportés au seul effet de l'intervention évaluée et ne sont pas en rapport avec l'instrument de mesure, à des biais dans le protocole de recherche ou à d'autres facteurs extérieurs. La validité interne signifie donc que le résultat de l'étude représente la réalité ;
- validité externe (*external validity*) : elle renseigne sur la « généralisabilité » des résultats de l'étude. Elle nécessite que le résultat soit confirmé par au moins une autre étude, ou par une méta-analyse sans hétérogénéité, ou bien par une cohérence avec les données fondamentales prouvées (biologiques, épidémiologiques, physiopathologiques) ;
- validité de critère (*criterion validity*) : appelée aussi validité prédictive, elle est évoquée lorsque les résultats obtenus par un instrument sont identiques aux résultats obtenus par un autre instrument dont la validité est déjà prouvée (instrument de référence ou *gold standard*).

VARIABILITÉ INTEROBSERVATEURS, INTRA OBSERVATEUR

Interobserver, intraobserver variability

La variabilité interobservateurs signifie la différence d'observation ou d'évaluation d'un événement ou d'une mesure par deux (ou plus) chercheurs différents. Elle est mesurée par le coefficient kappa.

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
P
q
r
s
t
u
v
w
x
y
z

a La variabilité intraobservateur signifie la différence d'observation
b ou d'évaluation d'un événement ou d'une mesure par le même
c chercheur, généralement à des moment différents. Elle peut être
d évaluée par des tests appariés.

d VARIABLE

e Variable

f La variable est le critère analysé dans les calculs statistiques. On
g distingue différents types de variables, catégorielle, censurée, dépendante,
h indépendante, métrique :

- i • variable catégorielle : elle peut être nominale ou ordinale. La
j variable nominale, appelée aussi binaire ou dichotomique, ne
k peut prendre que deux valeurs ou deux caractéristiques, par
l exemple décès oui/non, vomissements oui/non, femme/homme.
m Une variable catégorielle peut aussi avoir plusieurs valeurs telles
n que le groupe sanguin ou le mode de sortie de l'hôpital (décès,
o retour à domicile, ou maison de convalescence). La variable ordi-
p nale peut avoir ou non une suite hiérarchique ou une échelle. Un
q exemple de variable ordinale à suite hiérarchique est l'index de
r masse corporelle, pour définir : la maigreur, le poids normal, le
s surpoids, l'obésité, la superobésité ; un exemple de variable ordi-
t nale non hiérarchique est un infarctus du myocarde antérieur,
u antéroseptal, postérieur ;
- v • variable censurée (ou tronquée) : se dit d'une variable qui évolue
w dans le temps comme la survie ou les délais de survenue d'évé-
x nements comme la récurrence d'une maladie (cancers, maladies
y inflammatoires). Une variable censurée peut devenir catégorielle,
z par exemple récurrences avant ou après 5 ans ;
- variable dépendante : se dit d'une variable qui change (ou varie) en
relation avec une ou d'autres variables considérées dans l'étude ;
- variable indépendante : se dit d'une variable dont la valeur ou la
caractéristique ne dépend pas du résultat des autres variables de
l'étude (mais peut contribuer à la variance). Par exemple, dans la
relation entre « fumer » et le cancer du poumon, le cancer est la
variable dépendante et « fumer » est la variable indépendante ;
- variable métrique ou quantitative : la variable métrique est tou-
jours présentée sous une forme numérique. On distingue les varia-
bles « continues », représentées par un nombre infini de valeurs
(taille, poids, pression artérielle ou glycémie...) et les variables
« discontinues » (ou discrètes), représentées par des nombres
entiers (nombre de poussées d'une maladie, nombre de ganglions
dans une pièce de résection digestive, ou nombre d'enfants...).

VARIANCE (ANALYSE DE)

Analysis of variance, ANOVA

La variance est une mesure de dispersion de la distribution autour de la moyenne, dont la racine carrée est l'écart type. L'analyse de variance est un préliminaire indispensable à la comparaison de moyennes entre plusieurs échantillons indépendants, en supposant que les variables suivent une distribution normale. C'est une technique discriminante qui permet d'affirmer l'existence d'une différence sans pour autant la localiser. Elle peut donc faire appel ensuite à des tests de comparaison de moyennes entre plusieurs échantillons (MCP, pour *Multiple Comparison Tests*), choisis selon l'habitude et le contexte. Certains de ces tests sont volontairement libéraux (ils traquent la petite différence), d'autres sont au contraire conservatifs (ils ne gardent que les différences flagrantes) ; en médecine, la rigueur veut que l'on soit plutôt conservatif (notamment dans les essais thérapeutiques).

L'analyse de variance peut être simple (effet d'un seul facteur, par exemple, l'effet de la kinésithérapie sur les complications bronchopulmonaires postopératoires) ou multiple (MANOVA, par exemple, effet de plusieurs techniques de kinésithérapie).

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

W

WILCOXON (TEST DE)

→ TESTS STATISTIQUES

y

YATES (CORRECTION)

→ CORRECTION DE YATES

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

W

WILCOXON (TEST DE)

→ TESTS STATISTIQUES

y

YATES (CORRECTION)

→ CORRECTION DE YATES

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

Annexes

Annexe 1. Items du STARD (études évaluant les moyens diagnostiques)

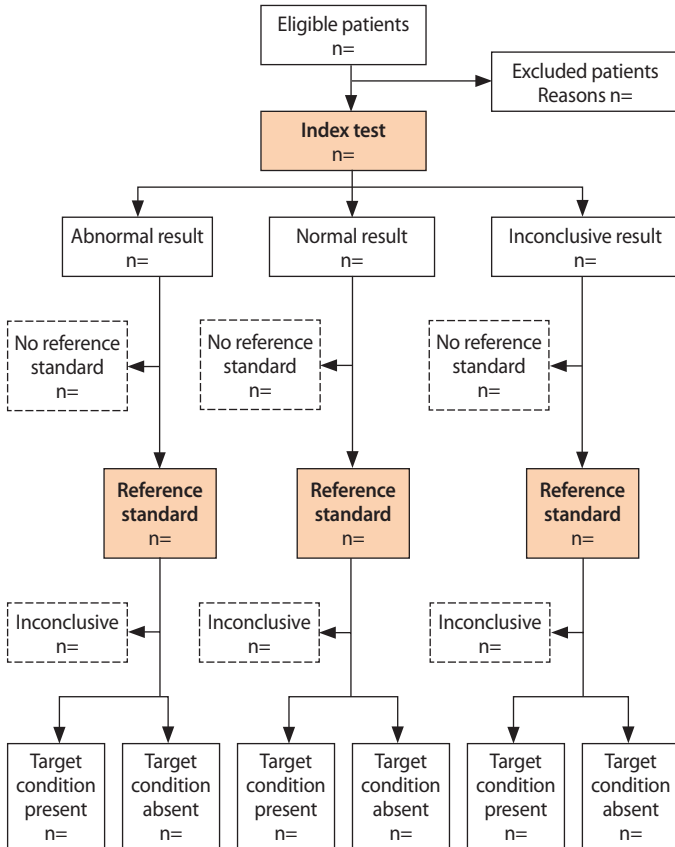
<i>Section and Topic</i>	<i>Item</i>		<i>On page</i>
Title/abstract/ keywords	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	
Introduction	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	
Methods			
Participants	3	Describe the study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	
	4	Describe participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	Describe participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	
	6	Describe data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	
Test methods	7	Describe the reference standard and its rationale.	
	8	Describe technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	

	9	Describe definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard.	
	10	Describe the number, training and expertise of the persons executing and reading the index tests and the reference standard.	
	11	Describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	
Statistical methods	12	Describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	
	13	Describe methods for calculating test reproducibility, if done.	
Results			
Participants	14	Report when study was done, including beginning and ending dates of recruitment.	
	15	Report clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	
	16	Report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).	
Test results	17	Report time interval from the index tests to the reference standard, and any treatment administered between.	

	18	Report distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	
	19	Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20	Report any adverse events from performing the index tests or the reference standard.	
Estimates	21	Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	
	22	Report how indeterminate results, missing responses and outliers of the index tests were handled.	
	23	Report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	
	24	Report estimates of test reproducibility, if done.	
Discussion	25	Discuss the clinical applicability of the study findings.	

(D'après Bossuyt P.M. *et al* Towards Complete and Accurate Reporting of Studies of Diagnostic Accuracy : The STARD Initiative. *Ann Intern Med* 2003 ; 138 : 40-4.)

Annexe 2. Diagramme du flux des études évaluant les moyens diagnostiques (STARD)



(D'après Bossuyt P.M. *et al* Towards Complete and Accurate Reporting of Studies of Diagnostic Accuracy : The STARD Initiative. *Ann Intern Med* 2003 ; 138 : 40-4.)

Annexe 3. Items du STROBE.

<i>Item</i>	<i>Item Number</i>	<i>Recommendation</i>
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract. (b) Provide in the abstract an informative and balanced summary of what was done and what was found.
Introduction		
Background/ Rationale	2	Explain the scientific background and rationale for the investigation being reported.
Objectives	3	State specific objectives, including any prespecified hypotheses.
Methods		
Study design	4	Present key elements of study design early in the paper.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.
Participants	6	(a) Cohort study: Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. Case-control study: Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. Cross-sectional study: Give the eligibility criteria, and the sources and methods of selection of participants. (b) Cohort study: For matched studies, give matching criteria and number of exposed and unexposed. Case-control study: For matched studies, give matching criteria and the number of controls per case.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.

Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.
Bias	9	Describe any efforts to address potential sources of bias.
Study size	10	Explain how the study size was arrived at.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why.
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding. (b) Describe any methods used to examine subgroups and interactions. (c) Explain how missing data were addressed. (d) Cohort study: If applicable, explain how loss to follow-up was addressed. Case-control study: If applicable, explain how matching of cases and controls was addressed. Cross-sectional study: If applicable, describe analytical methods taking account of sampling strategy. (e) Describe any sensitivity analyses.
Results		
Participants	13*	(a) Report the numbers of individuals at each stage of the study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analyzed. (b) Give reasons for nonparticipation at each stage. (c) Consider use of a flow diagram.
Descriptive data	14*	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders. (b) Indicate the number of participants with missing data for each variable of interest. (c) Cohort study: Summarize follow-up time—e.g., average and total amount.

Outcome data	15*	Cohort study: Report numbers of outcome events or summary measures over time. Case-control study: Report numbers in each exposure category or summary measures of exposure. Cross-sectional study: Report numbers of outcome events or summary measures.
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence intervals). Make clear which confounders were adjusted for and why they were included. (b) Report category boundaries when continuous variables were categorized. (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions and sensitivity analyses.
Discussion		
Key results	18	Summarize key results with reference to study objectives.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.
Generalizability	21	Discuss the generalizability (external validity) of the study results.
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.

*Give such information separately for cases and controls in case-control studies, and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

(D'après von Elm E. Altman D.G. Egger M. Pocock S.J. Gøtzsche P.C. *et al.* The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies, *PLoS Medicine* 2007 ; 4, No. 10, e296 doi:10.1371/journal.pmed.0040296.)

Liste des termes anglais

Actuarial (survival) :	actuarielle (survie)	a
Agreement testing :	concordance (test de)	b
Algorithm :	algorithme	c
Alpha risk :	alpha (risque)	d
ANOVA :	variance (analyse de)	e
Askmedline :	PICO	f
Bayesian analysis :	Bayes (analyse ou probabilité bayésienne)	g
Beta error :	bêta (risque)	h
Bias :	biais	i
Bland-Altman test :	Bland et Altman (test de)	j
Blind :	aveugle	k
Bonferroni correction :	correction dite de Bonferroni	l
Box plot :	boîte à moustache	m
Case series :	série de cas	n
Case-control study :	cas-témoin (étude)	o
Chi square :	Chi^2	p
Clinical relevance :	pertinence clinique	q
Clinical trial :	essai clinique	r
Cluster randomisation :	randomisation par grappe	s
Cochran Q test :	tests statistiques	t
Cochrane Collaboration :	Cochrane (Collaboration)	u
Cochrane Library :	Cochrane (Librairie)	v
Cochran-Mantel-Haenszel test :	tests statistiques	w
Cochran Q test :	tests statistiques	x
Coefficient of correlation :	corrélation (coefficient de)	y
Cohort study :	cohorte (étude de)	z
Concealment of allocation :	secret d'attribution	
Concordance testing :	concordance (test de)	
Confidence interval :	intervalle de confiance	
Confounding factor :	facteur de confusion	
CONSORT statement :	CONSORT	
Cost-benefit analysis :	coût-bénéfice (analyse)	
Cost-effectiveness analysis :	coût-efficacité (analyse)	
Cost-utility analysis :	coût-utilité (analyse)	
Cox model :	Cox (modèle de)	
Cross-over clinical trial :	essai clinique croisé	
Cross-sectional study :	étude transversale	
Delphi technique :	technique Delphes	

- a** Diagnostic accuracy : précision diagnostique
 Dichotomous variable : dichotomique (variable)
- b** Discrete variable : discontinue ou discrète (variable)
- c** Discriminant analysis : discriminante (analyse)
 Double blind : double aveugle
- d** Effect size (standardized) : ampleur de l'effet (standardisé)
 Efficiency : efficience
- e** Embase : embase
- f** Endpoint : critère de jugement
 Equivalence trial : essai d'équivalence
- g** Evidence-based medicine : factuelle (médecine)
 False negative : faux négatif
- h** False positive : faux positif
- i** Fisher exact test : tests statistiques
 Fixed effect : effet fixe
- j** Forest plot : forest plot ou graphique en forêt
 Friedman test : tests statistiques
- k** Funnel plot : funnel plot ou graphique en entonnoir
- l** Gaussian distribution : distribution normale
 Gold standard : gold standard
- m** Grades of recommendation : grades de recommandation
 Grey literature : grise (littérature)
- n** Guidelines for clinical practice : recommandations pour la pratique clinique
- o** Hawthorne effect : effet Hawthorne
- p** Hazards ratio : rapport de hasards
 Heterogeneity : hétérogénéité
- q** Homogeneity : homogénéité
 Incidence : incidence
- r** Inclusion criteria : critères d'inclusion
- s** Informed consent : consentement éclairé
 Intention to treat analysis : intention de traiter (analyse en)
- t** Inter or intraobservers variability : variabilité inter ou intraobservateurs
 Interquartile interval (or range) : intervalle interquartile
- u** Kaplan-Meier estimation : estimation de Kaplan-Meier
- v** Kappa value : kappa de Cohen
- w** Kruskal-Wallis test : tests statistiques
 Learning curve : courbe d'apprentissage
 Levels of evidence : niveaux de preuves
- x** Likelihood ratio : rapport de vraisemblance
- y** Linear regression : régression linéaire (analyse de)
 Logistic regression : régression logistique (analyse de)
- z** Mann-Whitney U test : tests statistiques

Mantel-Haenszel test : tests statistiques	a
Matching : appariement	b
McNemar test : tests statistiques	c
Mean difference : différence moyenne	d
Mean : moyenne	e
Median : médiane	f
Medline : Medline	g
MeSh : MeSh	h
Meta-analysis : méta-analyse	i
MOOSE : MOOSE	j
Multicentre study : étude multicentrique	k
Multiple linear regression : régression linéaire multiple (analyse de)	l
Non-inferiority trial : essai de non-infériorité	m
Normal (Gaussian) distribution : distribution normale (gaussienne)	n
Null hypothesis : hypothèse nulle	o
Number needed to harm : nombre nécessaire pour nuire	p
Number needed to treat : nombre de sujets à traiter	q
Odds ratio : cotes (rapport des)	r
Outcome : critère de jugement	s
Peto method : méthode de Peto	t
PICO : PICO	u
Placebo effect : effet placebo	v
Pooling : sommation	w
Predictive value : prédictive (valeur)	x
Prevalence : prévalence	y
Prevention : prévention	z
Primary sources : sources primaires	
Propensity score : propension (score de)	
Proportional hazards (Cox) model : Cox (modèle de)	
Prospective study : étude prospective	
Quality of life : qualité de vie	
QUOROM : QUOROM	
Random effect : effet aléatoire	
Random number tables : tables de nombres aléatoires	
Randomisation : randomisation	
Randomized controlled trial : essai contrôlé randomisé	
Relative risk : risque relatif	
Reliability : fiabilité	
Retrospective study : étude rétrospective	
Review : synthèse	
RevMan® : RevMan®	
Risk reduction : réduction de risque	
ROC curve : courbe ROC	

a Scale : échelle
Science citation index : science citation index

b Secondary sources : sources secondaires

c Sensitivity analysis : analyse de sensibilité
Sensitivity : sensibilité

d Specificity : spécificité
Standard deviation : écart type

e Standard error : erreur type

f Standardization : standardisation
STARD : STARD

g Statistical tests : tests statistiques
Statistically significant : statistiquement significatif

h STROBE : STROBE

i Stratification : stratification
Student test : tests statistiques

j Subgroups (analysis) : sous-groupes (analyse de ou en)
Survival : survie

k Systematic review : synthèse méthodique

l TREND : TREND
Two sided analysis : bilatéral (test)

m Two tailed analysis : bilatéral (test)
Type 1 error : alpha (risque)

n Type 2 error : bêta (risque)
Validity : validité

o Variable : variable

p Visual analogue scale : échelle visuelle analogique
Weighting : pondération

q Wilcoxon rank sum test : tests statistiques

r Yates correction : correction de Yates

s

t

u

v

w

x

y

z