

Notions

L'aveugle

L'objectif de faire une étude en aveugle est d'éviter les biais dus au fait que le participant ou le chercheur soit au courant de son statut. Un participant qui sait qu'il prend un placebo sera moins enclin à rester dans l'étude ou à observer les prises prévues. Un chercheur qui connaît le statut du participant peut changer son comportement lors d'une interaction, ou être moins objectif lors du traitement des données. L'aveugle peut être utilisé en simple, double ou triple. En simple seuls les participants ou les chercheurs ne seront pas au courant des statuts, en double les deux seront ignorant des statuts, et en triple les sujets, les intervenants directs et les analystes de données seront dans l'ignorance. En anglais on utilise les termes de *blinding* ou *masking*. Ainsi, une étude pourra être décrite comme une *double blinded study*.

Le bras

Ce terme, comme pour un fleuve, décrit l'organisation des groupes selon le traitement. On peut avoir deux, trois, ou quatre bras (rarement plus).

Expérimental – le groupe qui subit l'intervention

Comparaison active – lorsqu'on fait une étude pour établir la non-infériorité d'un traitement

Placebo – le produit qui ressemble à ce qu'on teste, mais sans substance active

Factice – (*sham* en anglais) équivalent du placebo dans un contexte non pharmaceutique (par exemple : un appareil inefficace, ou même une intervention chirurgicale sans objet)

Sans intervention – lorsqu'on compare un traitement avec l'absence de traitement

Attribution

La notion d'attribution fait référence à la manière dont on affecte les participants dans chacun des bras de l'étude. Il est possible d'utiliser une distribution aléatoire ou non aléatoire. Dans certains contextes la randomisation n'est pas possible pour une raison ou une autre, mais cela ne dédouane pas les auteurs de leur responsabilité de s'assurer de la comparabilité des groupes, et on devrait trouver dans la publication un tableau ou des données qui permettent de vérifier ce point. La randomisation permet d'obtenir des groupes comparables plus facilement, et l'étude randomisée se trouve au-dessus de l'étude non-randomisée dans la hiérarchie des niveaux de preuves.

On procédera à la randomisation a priori lorsqu'on a déjà recruté les sujets, et cela s'effectue soit par blocs permutés où un certain nombre de participants seront affectés au groupe A, puis autant au groupe B et ainsi de suite, soit par strates où on affectera de façon aléatoire, mais en tenant compte des caractéristiques des participants.

On peut également randomiser *a posteriori* pendant le processus de recrutement en fonction des attributions précédentes, parfois par strates. Dans le cas d'une étude longitudinale où on commence à avoir des éléments sur l'efficacité d'une intervention, on peut recourir à une méthodologie dite du gagnant renforcé.

Il y a de nombreuses ressources sur internet qui expliquent la randomisation en plus de détail. Pour ceux qui souhaitent en savoir plus sur le côté statistique vous pouvez lire l'article ci-dessous :

Lim, C.Y. and In, J., 2019. Randomization in clinical studies. *Korean journal of anesthesiology*, 72(3), p.221.

Cluster randomisation

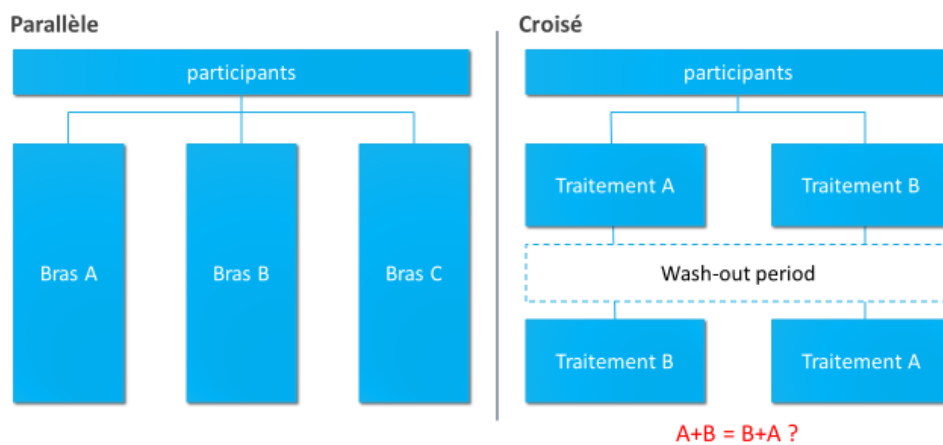
Dans certains contextes il est préférable d'éviter le risque qu'un membre d'un groupe influence un participant dans l'autre. Cette approche va randomiser par groupe plutôt qu'au niveau de l'individu. On l'utilise fréquemment d'évaluer une intervention où l'impact concerne le système plutôt que l'individu

Exemple : Essai de deux stratégies de perte de poids



Croisement

Dans une situation classique les bras sont organisés en parallèle, comme on voit sur la gauche dans l'illustration ci-dessous. Parfois on peut décider de procéder par le croisement. Cette approche a l'avantage de neutraliser une éventuelle défaillance de comparabilité entre les groupes. En principe on retrouve cette méthodologie le plus fréquemment dans les études qui comparent deux traitements. On voit à droite sur l'illustration que les traitements se croisent. Il y a toujours une période dite de *wash-out*, qui permet de s'assurer que l'on ne puisse pas imputer un effet d'un traitement à la présence continue du traitement précédent dans l'organisme.



Vérification de résultats

Il est important d'avoir confiance dans la fiabilité de ce qui est mesuré. A ce titre, lorsqu'on lit la partie méthodologie d'un article on va vérifier que les prises de mesures sont justes. On s'assurera que tout appareil est calibré, et que tout questionnaire ou autre outil est valide, c'est-à-dire qu'il a déjà été utilisé dans des études précédentes.

On peut effectuer un certain nombre de démarches en fonction de la situation. Parfois on aura besoin de procéder à une vérification en effectuant un même test après un laps de temps donné. On appelle cette procédure *test-retest*.

Dans les situations où il y a plusieurs personnes qui évaluent les sujets, il convient de vérifier ce qu'on nomme *inter-rater reliability*. Cela peut s'exprimer par un pourcentage qui décrit le degré d'accord entre évaluateurs, mais également par un score kappa qui varie de -1 à 1. Les explications statistiques

et les mérites et inconvénients des approches sont évoqués dans un article de 2012ⁱ. Pour le simple lecteur d'article, on peut se contenter de vérifier que les auteurs ont bel et bien procédé à cette démarche si la situation s'y prête.

Spécificité et sensibilité

Ces deux mesures permettent d'évaluer à quel point un test mesure ce qu'il est censé mesurer, et on l'utilise notamment pour des tests biologiques au laboratoire, mais on le trouvera aussi dans d'autres contextes.

La **spécificité** s'assure de la bonne identification des vrais négatifs, avec peu de faux **positifs**, alors que la **sensibilité** va évaluer l'identification correcte des vrais positifs, avec un minimum de faux **négatifs**.

On l'exprime par un pourcentage, et le schéma ci-dessous montre comment on obtient ce score, ainsi que PPV et NPV (*positive predictive value* et *negative predictive value*, ou la capacité de prédire correctement des cas positifs et négatifs respectivement)

		The Truth		
		Has the disease	Does not have the disease	
Test Score:	Positive	True Positives (TP) a	False Positives (FP) b	$PPV = \frac{TP}{TP + FP}$
	Negative	False Negatives (FN) c	True Negatives (TN) d	
		Sensitivity $\frac{TP}{TP + FN}$	Specificity $\frac{TN}{TN + FP}$	
Or,		$\frac{a}{a + c}$	$\frac{d}{d + b}$	

Vérification des statistiques

Il est important aussi de vérifier les résultats de l'analyse statistique.

En effet, certaines méthodes statistiques engendrent des erreurs de type I, où il y a le rejet de l'hypothèse nulle vraie, ou de type II où on ne rejette pas une hypothèse nulle fausse.



Image source: unbiasedresearch.blogspot.com

Souvent il s'agit de méthodes statistiques où il y a de nombreuses variables ou des mesures répétées qui donnent lieu à ces erreurs. Il convient de vérifier que les auteurs ont procédé aux **tests post hoc** nécessaires. On trouvera souvent le terme *correction* ou *post hoc*.

De même, il peut être nécessaire de recourir à une **stratification**, notamment lorsque l'échantillon est grand, et ceci nous permet d'avoir des résultats affinés par rapport à un chiffre global brut. Dans les situations où on redoute la présence de facteurs de confusion, on peut procéder à des **ajustements** qui permettent d'avoir une idée plus claire de quelles variables sont réellement responsables d'un effet statistique.

L'essai randomisé contrôlé

Ce type d'étude est considéré comme une référence dans la recherche scientifique, et se situe sur le rang le plus élevé de la pyramide des preuves. On appelle ce statut *gold standard* en anglais. Souvent trouvé avec l'abréviation RCT : *randomized controled trial* (en américain) ou *randomised controlled trial* (anglais), c'est une méthodologie rigoureuse pour identifier une relation cause-effet entre un traitement et un résultat.

On choisira le type d'attribution au besoin, et on utilisera l'aveugle lorsque c'est possible.

On voit dans le tableau ci-dessous qu'il présente de nombreux avantages, mais n'est pas 100% parfait.

+	-
<ul style="list-style-type: none">• Niveau de preuve le plus élevé pour évaluer un traitement• Meilleure démonstration de lien cause à effet• La randomisation limite l'influence des facteurs de confusion potentiels• Pas de doute sur la chronologie• Permet les études en aveugle et ainsi limite certains biais• Peut mesurer incidence et résultats multiples	<ul style="list-style-type: none">• Problèmes éthiques pour l'attribution.• Coûteux et long.• Peu efficace lorsque la pathologie est rare ou lorsque l'effet est retardé.• Generalisabilité – les participants peuvent être plus enclins à adhérer au traitement que la population plus large.

Dans un article de 2014, Gamble plaide pour la complémentarité des études. En effet, le fait d'envisager la recherche en termes d'hierarchie peut peut-être pousser une équipe vers un essai randomisé alors que la question de recherche pourrait être traitée par une étude de cohorte ou une étude cas-témoin.

Dans un tableau (reproduit ci-dessous) il tente de démontrer comment les limites des études randomisées peuvent être compensées par une méthodologie dite inférieure.

Limite de l'essai randomisé	Complémentarité des études de cohorte et cas-témoin
Protocole stricte et peu représentatif d'une situation empirique	Habituellement représentatif des soins médicaux réels
Exclut des populations d'intérêt (enfants, femmes enceintes, personnes âgées)	Peut focaliser sur tout type de population
Taille d'échantillon limitée	Peut inclure un grand nombre de patients, surtout en utilisant des sources de données secondaires, ce qui est probant pour la détection d'événements rares.
Durée courte	Peut suivre une population pendant des années
Evaluation de comparaisons de traitements sans pertinence	Peut comparer plusieurs traitements pertinents
Critères d'évaluation pas toujours pertinents pour le patient	Peut inclure tout critère qui est mesurable
Coût élevé	Coût relativement bas.

Source : John-Michael Gamble, Canadian Journal of Hospital Pharmacy , VOLUME 67 , NUMBER 5 , September-October 2014

Modèles d'analyse

Specialist Unit for Review Evidence (SURE) 2016.

Ce modèleⁱⁱ a été conçu à Cardiff, et se base sur une liste de questions

1. Does the study address a clearly focused question/hypothesis?
Population/Problem?
Intervention?
Comparator/control?
Outcomes?
Can you identify the primary outcome?

2. Was the population randomised?
If YES, were appropriate methods used?
e.g.: random number tables, opaque envelopes

Note: The following methods are not appropriate: alternating participants coin toss, birth dates, record numbers, days of the week

3. Was allocation to intervention or comparator groups concealed?
Is it possible for those allocating to know which group they are allocating people to?

As above, methods such as alternating participants coin toss, birth dates, record numbers, days of the week will not allow appropriate allocation concealment.

4. Were participants/investigators blinded to group allocation?
If NO, was assessment of outcomes blinded?
5. Were interventions (and comparisons) well described and appropriate?
Aside from the intervention, were the groups treated equally?
Was exposure to intervention and comparison adequate?
Was contamination acceptably low?

6. Was ethical approval sought and received?
Do the authors report this?
7. Was a trial protocol published?
Was a protocol published in a journal or clinical trial registry before participants were recruited?
If a protocol is available, are the outcomes reported in the paper listed in the protocol?
8. Were the groups similar at the start of the trial?
Are baseline characteristics provided and discussed (e.g. age, sex, social class, lifestyle etc.)?
Are any differences >10%?
9. Was the sample size sufficient?
Were there enough participants?
Was there a power calculation? If YES, for which outcome?
Were there sufficient participants for the whole duration?
10. Were participants properly accounted for?
Was follow-up $\geq 80\%$?
Were patients analysed in the groups to which they were randomised?
Was an Intention to Treat analysis conducted?
Was the follow-up period long enough?
11. Data analysis
Are the statistical methods well described?
Consider: How missing data was handled; were potential sources of bias (confounding factors) controlled for; How loss to follow-up was addressed.
12. Results
Were outcome measures reliable (e.g. objective or subjective measures)?
Were all outcome measurements complete?
Were all important outcomes assessed?
Are the authors' conclusions adequately supported by the results?
13. Is any sponsorship/conflict of interest reported?
14. Finally...consider:
Did the authors identify any limitations?
Are the conclusions the same in the abstract and the full text?

Quality control

An entire branch of research has developed around how research is carried out and reported. It carries several names; metascience, meta-research or evidence-based research. Its main development began after a paper was published in JAMA in 1966ⁱⁱⁱ in which it was found that 73% of papers contained erroneous conclusions drawn from the data. One of the most prolific researchers in the field is John P. A. Ioannidis, and many of his studies have been of note. The field contains five subdomains:

1. **methods**, which deals in particular with statistical methods
2. **reporting**, concerning the publication of studies
3. **reproducibility**, a major problem in research
4. **evaluation**, and how to improve the peer-review process
5. **incentives**, or how to ensure that researchers avoid deontological slips due to the "publish or perish" mindset

Reporting guidelines have become standard practice since they were first introduced in 1993 as an attempt to ensure that a number of important elements were systematically included in any published trial. The original 32 item checklist and flow diagram has been revised at regular intervals since then, and now bears the name CONSORT. Most well-known journals require any randomised trial to use this tool. Since then, other groups were created to produce tools for other types of studies, such as TREND for non-randomised trials, PRISMA for systematic reviews and meta-analyses, and STROBE for observational studies...

While these tools were created with the authors in mind, a reader can obtain a rapid grasp of how well a paper is constructed based on the degree to which the study adheres to the checklist, and whether the flow diagram, when appropriate, is used.

Travail personnel

Visitez le site <https://www.equator-network.org/> et trouvez les documents CONSORT concernant l'essai randomisé. A partir de cette liste, et de l'outil SURE (ci-dessus) faites une ébauche de commentaire de l'article cité ci-dessous.

Zhang, J., Brackbill, D., Yang, S., Becker, J., Herbert, N. and Centola, D., 2016. Support or competition? How online social networks increase physical activity: a randomized controlled trial. *Preventive medicine reports*, 4, pp.453-458.

Vous pouvez travailler en groupe, et me contacter avec vos questions.

Un commentaire sera publié sur mon site avant la fin de la semaine.

ⁱ McHugh, M.L., 2012. Interrater reliability: the kappa statistic. *Biochimica medica: Biochimica medica*, 22(3), pp.276-282.

ⁱⁱ Specialist Unit for Review Evidence (SURE) 2016.

Questions to assist with the critical appraisal of randomised controlled trials and other experimental studies available at: <http://www.cardiff.ac.uk/specialist-unit-for-review-evidence/resources/critical-appraisal-checklists>

ⁱⁱⁱ Schor, Stanley (1966). "Statistical Evaluation of Medical Journal Manuscripts". *JAMA: The Journal of the American Medical Association*. **195** (13): 1123. [doi:10.1001/jama.1966.03100130097026](https://doi.org/10.1001/jama.1966.03100130097026)