Étude randomisée en double aveugle

L'**étude randomisée en double aveugle**, *avec répartition aléatoire*, *randomisée* ou **à double insu** (ou *en double aveugle*) est une démarche expérimentale utilisée dans de nombreuses disciplines de recherche telles que la médecine, les sciences sociales et la psychologie, les sciences naturelles telles que la physique et la biologie.

En pharmacie, elle est utilisée dans le développement de nouveaux médicaments et pour évaluer l'efficacité d'une démarche ou d'un traitement. Par exemple, durant l'étude sur un médicament, ni le patient ni le prescripteur ne savent si le patient utilise le médicament actif ou le placebo. Le rôle d'un tel protocole, relativement lourd à mettre en place, est de réduire au mieux l'influence sur la ou les variables mesurées que pourrait avoir la connaissance d'une information à la fois sur le patient (premier « aveugle ») et sur le médecin (deuxième « aveugle »). C'est la base de la médecine fondée sur les faits.

Il existe même un raffinement supplémentaire avec la procédure dite en « triple aveugle » ou « triple insu », où le troisième « aveugle » est le chercheur qui opère le traitement statistique des données. Bien que moins courante, cette démarche offre un niveau de preuve encore amélioré[1].

Sommaire

Histoire

L'utilisation des statistiques pour montrer l'efficacité d'un traitement remonte au XIX^e siècle : le physicien Pierre-Charles Alexandre Louis (1787-1872) montra que le traitement de la pneumonie par des sangsues n'était pas bénéfique mais délétère[2].

Les premières études en « simple aveugle », le patient ignorant s'il reçoit le vrai traitement ou un placebo, apparaissent dès la fin du XIX^e siècle pour prouver la supercherie du magnétisme animal développé par Franz-Anton Mesmer, ainsi que d'autres techniques « magnétiques ». Armand Trousseau (1801-1867) invente les premières pilules placebos, faites à base de mie de pain et démontre ainsi leur équivalence au niveau efficacité avec les médicaments homéopathiques[3].

Description : l'exemple de la recherche médicale

Un des problèmes de la recherche médicale est que l'on ne peut pas faire varier un paramètre en laissant les autres constants : la vie est constituée d'un équilibre et la variation d'un paramètre a une répercussion sur les autres (réaction d'équilibrage de l'organisme, homéostasie). Un autre problème est que les personnes réagissent de manières très différentes, et que la réaction d'une personne peut varier selon le moment où est faite l'étude ; certaines personnes guérissent spontanément d'une maladie, d'autres réagissent plus ou moins bien aux médicaments, et par ailleurs, le fait même de prendre un traitement peut parfois avoir des effets bénéfiques ou négatifs même si le traitement lui-même est sans effet (effet placebo). L'idée est de réduire l'influence de la subjectivité des intervenants.

Comme il est impossible de s'affranchir de la diversité humaine, il faut la prendre en compte dans l'étude. On constitue donc deux groupes de patients, l'un prenant un traitement contenant le principe actif (le médicament), l'autre prenant un placebo (traitement sans principe actif, présenté généralement sous la même forme galénique). La répartition principe actif/placebo se fait de manière aléatoire et ni la personne prenant le traitement, ni la personne l'administrant ne savent s'il y a du principe actif (*double insu*). La levée du voile n'est faite qu'après le traitement statistique.

On ne pourra dire qu'un traitement a de l'effet que si l'on observe une différence statistique *significative* entre les deux groupes, c'est-à-dire que la probabilité que la différence observée entre les deux traitements soit due uniquement au hasard, est inférieure à un certain seuil fixé. En médecine, ce seuil est très souvent fixé à 5 %.

Traitement statistique

Aucun traitement n'est efficace à 100 % (sur tout le monde dans tous les cas). La guérison naturelle n'est pas non plus systématique. Il faut donc étudier un nombre de cas suffisamment important pour pouvoir écarter un biais statistique.

Cas binaire

Plaçons-nous dans un cas « binaire » : la personne guérit ou ne guérit pas. Nous avons donc deux groupes, le groupe « m » qui a reçu le médicament et le groupe « p » qui a reçu le placebo.

Groupes de taille identique

On supposera que chaque groupe comprend n personnes (ils sont de même taille).

Dans le groupe « m », le nombre de personnes ayant guéri est O_m (O pour « observé »). Dans le groupe « p », le nombre de personnes ayant guéri est O_p. Les taux de guérison respectifs p_m et p_p sont donc :

$$p_\mathrm{m} = O_\mathrm{m} / n$$
$$p_\mathrm{p} = O_\mathrm{p} / n$$

Le tableau de résultat est :

Résultats de l'essai en double insu

	Groupe « m »	**Groupe « p »**
Guéris	O_m	O_p
Non guéris	$n - O_m$	$n - O_p$

On utilise un test du χ^2 d'indépendance, ou test du χ^2 de Pearson : on a deux hypothèses

- l'hypothèse dite « nulle », H_0 : le médicament n'a aucun effet, le taux de guérison est le même dans les deux groupes, l'écart constaté est dû à des fluctuations statistiques ;
- l'hypothèse dite « alternative », H_1 : le médicament est efficace, le taux de guérison est supérieur dans le groupe « m » par rapport au groupe « p ».

Selon l'hypothèse nulle, on peut fusionner les deux groupes. On a donc un groupe de $2\times n$ personnes, et un nombre de guérisons égal à $O_m + O_p$. Le taux de guérison p_0 dans l'hypothèse nulle est donc

$$p_0 = (O_m + O_p)/(2\times n)$$

Donc, dans l'hypothèse H_0, le nombre de guérisons dans le groupe « m » comme dans le groupe « p » devrait être E (E pour « espéré », ou *« expected »* en anglais) :

$$E = p_0 \times n$$

On devrait donc avoir le tableau suivant.

Résultats théoriques sous l'hypothèse nulle

	Groupe « m »	**Groupe « p »**
Guéris	E	E
Non guéris	$n - E$	$n - E$

Le χ^2 est la somme, pour toutes les cases du tableau, des différences au carré entre la valeur théorique et la valeur observée, divisées par la valeur théorique :

$$\chi^2 = \frac{(E - O_m)^2}{E} + \frac{(E - O_p)^2}{E} + \frac{(n - E - (n - O_m))^2}{E} + \frac{(n - E - (n - O_p))^2}{E}$$

soit en l'occurrence

$$\chi^2 = \frac{2 \times (E - O_m)^2 + 2 \times (E - O_p)^2}{E}$$

Il faut comparer cette valeur à la valeur tabulée, en considérant un risque d'erreur, typiquement 5 %, et le nombre de degrés de liberté, qui est le produit

(nombre de lignes du tableau - 1)×(nombre de colonnes du tableau - 1),

soit 1 degré de liberté ici. On se place dans le cas d'un test bilatéral, c'est-à-dire que l'on cherche juste à savoir si les valeurs sont différentes ou pas, sans préjuger du sens de la différence.

Loi du χ^2 à un degré de liberté pour un test bilatéral

Erreur admissible (p)	**50 % (p = 0,5)**	**10 % (p = 0,1)**	**5 % (p = 0,05)**	**2,5 % (p = 0,025)**	**1 % (p = 0,01)**	**0,1 % (p = 0,001)**
χ^2	0,45	2,71	3,84	5,02	6,63	10,83

Donc, pour un risque d'erreur de 5 % :

- si $\chi^2 \leq 3{,}84$, l'hypothèse H_0 est acceptée, on estime que le médicament n'a pas d'effet propre ;
- si $\chi^2 > 3{,}84$, l'hypothèse H_0 est rejetée, on estime que le médicament a un effet propre (qui peut être bénéfique ou délétère…).

Groupes de tailles différentes

On a maintenant un groupe « m » de taille n_m avec O_m guérisons, et un groupe « p » de taille n_p avec O_p guérisons. Le tableau des valeurs observées est :

Résultats de l'essai en double insu

	Groupe « m »	**Groupe « p »**
Guéris	O_m	O_p
Non guéris	n_m-O_m	n_p-O_p

On a

$$p_0 = (O_m + O_p)/(n_m + n_p).$$

Dans l'hypothèse H_0, le nombre de guérisons dans le groupe « m » devrait être E_m et le nombre de guérisons dans le groupe « p » devrait être E_p :

$$E_m = p_0 \times n_m$$
$$E_p = p_0 \times n_p$$

On devrait donc avoir le tableau suivant.

Résultats théoriques sous l'hypothèse nulle

	Groupe « m »	**Groupe « p »**
Guéris	E_m	E_p
Non guéris	n_m-E_m	n_p-E_p

Le χ^2 est :

$$\chi^2 = \frac{(E_m - O_m)^2}{E_m} + \frac{(E_p - O_p)^2}{E_p} + \frac{(n_m - E_m - (n_m - O_m))^2}{E_m} + \frac{(n_p - E_p - (n_p - O_p))^2}{E_p}$$

soit

$$\chi^2 = 2 \times \frac{(E_m - O_m)^2}{E_m} + 2 \times \frac{(E_p - O_p)^2}{E_p}.$$

On compare de même cette valeur à la valeur tabulée pour valider ou invalider l'hypothèse nulle.

Exemple

Le groupe « m » a 98 personnes et 19 ont guéri ; le groupe « p » a 101 personnes et 8 ont guéri. On a donc les tableaux suivants :

Résultats de l'essai en double insu

	Groupe « m »	Groupe « p »
Guéris	19	8
Non guéris	79	93

La probabilité dans l'hypothèse nulle est

$p_0 = (19 + 8)/(98 + 101) \cong 0{,}136$

et les espérances sont

$E_m \cong 13{,}3$
$E_p \cong 13{,}7$

le χ^2 vaut donc

$$\chi^2 = 2 \times \frac{(13,3 - 19)^2}{13,3} + 2 \times \frac{(13,7 - 8)^2}{13,7} \simeq 9,6$$

l'hypothèse nulle est donc rejetée avec un risque d'erreur inférieur à 1 % (p = 0,01), puisque $\chi^2 > 6{,}63$; on peut estimer que le médicament est efficace.

Nombre de sujets nécessaires

Selon la règle classique, les effectifs théoriques E_i doivent être supérieurs ou égaux à 5 (cf. *Test du χ^2 > Conditions du test*). Cela signifie qu'il faut au moins vingt personnes, puisque l'on a quatre classes. Il en faut en fait plus puisque les fréquences sont rarement égales à 0,5.

Si p est la probabilité de l'événement auquel on s'intéresse et n la taille de la population étudiée, alors on estime que l'on doit avoir :

$n \times p \geq 5$ et $n \times (1 - p) \geq 5$

puisque $(1 - p)$ est la fréquence de l'événement complémentaire, soit :

$n \geq 5/p$ et $n \geq 5/(1 - p)$

Paramètre chiffré

Dans certains cas, l'étude ne classe pas les patients dans des groupes guéris/non-guéris, mais mesure un paramètre chiffrable, par exemple la durée de la maladie (en jours), le taux de telle ou telle substance, la valeur de tel paramètre physiologique (par exemple fraction d'éjection ventriculaire gauche, glycémie, …). Cette quantification — ou numérisation — de la maladie est parfois difficile à faire, par exemple dans le cas de la douleur, de la dépression.

Dans ce cas-là, le paramètre est évalué patient par patient. Il en résulte deux ensembles de valeurs, un pour le groupe « m » et un pour le groupe « p ». Ces ensembles de valeurs sont en général résumés par deux valeurs, la moyenne E_i et l'écart type σ_i :

- la moyenne représente la tendance générale du groupe i ;
- l'écart type représente l'étalement des valeurs.

La première question à se poser est la loi que suivent les valeurs au sein d'un groupe. La plupart du temps, on estime qu'elles suivent une loi normale, mais il faut penser à le vérifier.

On détermine ensuite les intervalles de confiance : pour chacun des groupes, on détermine les valeurs entre lesquelles la moyenne se situerait avec 95 %, ou 99 % de probabilité, si l'on avait accès à une infinité de patients. On utilise pour cela la loi de Student : l'intervalle de confiance est de la forme

$$[E - t_{\gamma}^{ni-1} \cdot \sigma \ ; E + t_{\gamma}^{ni-1} \cdot \sigma]$$

où t_{γ}^{ni-1} est le quantile de la loi de Student pour :

- n_i -1 degrés de liberté, n_i étant l'effectif du groupe i ;
- γ est lié au niveau de confiance α par la formule suivante :
  γ = (1-α)/2.

Pour que l'on puisse distinguer les deux groupes, il faut que les espérances E_{m} et E_{p} soient suffisamment éloignées pour ne pas figurer dans l'intervalle de confiance de l'autre groupe.

Facteur de Student *t*

Effectif (*n*)	**Niveau de confiance (α)**				
	50 % (α = 0,5 ; γ = 0,25)	**90 % (α = 0,9 ; γ = 0,05)**	**95 % (α = 0,95 ; γ = 0,025)**	**99 % (α = 0,99 ; γ = 0,005)**	**99,9 % (α = 0,999 ; γ = 0,000 5)**
5	0,741	2,132	2,776	4,604	8,610
10	0,703	1,833	2,262	3,250	4,781
20	0,688	1,729	2,093	2,861	3,883
50	0,679	1,676	2,009	2,678	3,496
100	0,677	1,660	1,984	2,626	3,390
∞	0,674	1,645	1,960	2,576	3,291

Exemple

Dans le groupe placebo, on détermine que la maladie guérit en moyenne en dix jours, avec un écart type de 2 jours. Si le groupe comprend 100 personnes, alors l'intervalle de confiance pour un niveau de confiance de 95 % est de 6–14 jours (10 - 2×1,984 ≅ 6, 10 + 2×1,984 ≅ 14). Si la moyenne de la durée de la maladie avec le groupe médicament est inférieure à 6, on peut alors dire que le médicament est efficace avec un risque de 2,5 % d'erreur (puisque les 5 % de cas résiduels du groupe placebo sont répartis de part et d'autre de la moyenne, il y en a 2,5 % en dessous de 6). Il faudrait que la moyenne du groupe médicament soit inférieure à 3 jours pour être sûr avec 0,5 % d'erreur (10 - 2×3,390 ≅ 3).

Pertinence du test

Risque alpha : risque de faux négatif.

Risque bêta : puissance du test (sélectivité entre les deux populations).

Autres champs d'application

Le test en double insu s'applique aussi lorsque l'on veut tester l'efficacité d'un nouveau traitement par rapport à un autre, ce dernier étant alors appelé « traitement de référence » : il s'agit de déterminer si le nouveau traitement proposé est *significativement* plus efficace que l'ancien.

Le test en double insu est également couramment utilisé en dehors du domaine médical dès lors que l'on souhaite réaliser une étude s'affranchissant des biais de perceptions conscients ou non du sujet testé (préjugés). C'est notamment le cas lors d'études comparatives en marketing ou pour des tests organoleptiques (mesure de la qualité gustative d'un aliment par un jury).

Surveillance

Dans le cadre de la recherche biomédicale, une étude en double aveugle peut amener des difficultés dans la détection d'une sous-efficacité du traitement testé par rapport à son comparateur, ou d'une fréquence plus élevée des effets indésirables (l'aveugle empêchant de savoir si un groupe de traitement est significativement différent avant la levée d'insu). Dans des essais cliniques où la surveillance apparaît critique (essais multicentriques multinationaux par exemple, où la surveillance peut être complexe), il peut être décidé de constituer un Comité de Surveillance et de Suivi, indépendant du promoteur, qui examine les données (sans levée d'insu) lors d'analyses intermédiaires.

Problèmes éthiques

Si l'efficacité statistique des études en double aveugle ne fait aucun doute, celles-ci ne sont pas sans poser des problèmes éthiques. En effet, si les essais mettent en jeu des cobayes animaux, s'ils font intervenir des pathologies bénignes, ou des traitements d'une efficacité présumée douteuse (homéopathie), il n'y a alors pas de problème pour employer des placebos. La situation est cependant différente dans le cas de pathologies lourdes, qui font intervenir des tests sur des humains[4]. Fournir des placebos à des patients, alors que des études pilotes ont pu par exemple par ailleurs confirmer l'efficacité d'un traitement, frôle avec les limites de l'éthique médicale et des principes annoncés par la Déclaration d'Helsinki. Les responsables des études ont en effet souvent tendance à dégager leur responsabilité par la signature du consentement éclairé du patient, alors que le plus souvent ce dernier n'a pas d'autre alternative que de donner son accord sinon de refuser tout traitement, en étant peu à même de juger de la pertinence ou de l'opportunité des essais auxquels il est soumis.

Le recours à des études en double aveugle, et donc à l'utilisation de placebos, se fait parfois de façon assez automatique, sinon systématique comme dans le cas de la FDA, alors que l'on pourrait parfois se contenter de comparer l'efficacité de deux traitements, ayant validé au moins la Phase 2 d'une étude clinique, sans employer de groupe de contrôle, de vérifier l'absence de biais de sélection, ceci pour le bénéfice des patients[5,6]. Aucune réglementation en ce sens n'a cependant été adoptée en France en 2020, qui permette d'encadrer les avis des comités de protection des personnes.

Notes et références

1. Pierre Chevalier, *Démarche clinique et médecine factuelle*, Presses universitaires de Louvain, 12 février 2015 (ISBN 978-2-87558-374-1, lire en ligne (https://books.google.fr/books?id=zj8wBwAAQBAJ&q=aveugle+alias+insu+single+blind))
2. Le Quotidien du Médecin : toute l'information et la formation médicale continue des médecins généralistes et spécialistes (http://www.quotimed.com/journal/index.cfm?fuseaction=viewarticle&DArtIdx=360822)
3. Chamayou G, *L'essai « contre placebo » et le charlatanisme*, Les génies de la science, février-avril 2009, p. 14-17
4. « L'éthique de la recherche et l'usage du placebo : un état de la question au Canada », *Site Medecinesscience.org*, 14 janvier 2004 (lire en ligne (https://www.medecinesciences.org/en/articles/medsci/full_html/2004/01/medsci2004201p118/medsci2004201p118.html))
5. « Les aspects éthiques du placebo », *unblog*, 14 janvier 2004 (lire en ligne (http://leplacebo.unblog.fr/les-problemes-lies-a-l%E2%80%99utilisation-du-placebo-2/b-les-aspects-ethiques/))
6. « Essais cliniques : théorie, pratique et critique (4e ed.), Chap XV, les essais cliniques randomisés », *Lavoisier*, 2006

Voir aussi

Articles connexes

- Bonne pratique clinique
- Essai clinique
- Essai randomisé contrôlé

Liens externes

- Méthode et statistiques en médecine (http://www.med.univ-rennes1.fr/etud/pharmaco/methode_et_statistiques.htm), Stéphane Schück, Université de Rennes
- Un problème d'effectif (http://www.opimed.org/article.php?id_article=37), Opimed

---