

## Le biais d'observation : importance du triple aveugle

Pierre Chevalier  
Centre Académique de Médecine  
Générale, Université Catholique de  
Louvain

Texte sous la  
responsabilité de la  
rédaction franco-  
phone

### Biais d'observation

- Synonymes : biais de détection ou biais de constatation
- Anglais : observer bias, ascertainment bias, detection bias, interviewer bias

Dans une RCT, la non connaissance du traitement réellement reçu est appelée aveugle (alias insu). Cet insu peut se situer à différents niveaux : participants, soignants, évaluateurs des résultats, analystes des résultats, rédacteurs du manuscrit (Cochrane Handbook<sup>1</sup>). Quand il s'agit des participants et des soignants qui sont en insu, nous parlons de double aveugle. Le non respect de ce double aveugle peut biaiser les résultats de l'étude<sup>2,3</sup>.

Quand un double aveugle n'est pas possible, il reste la possibilité de faire l'évaluation en insu : la personne qui évalue les résultats ne sait pas dans quel bras d'étude est inclus le patient dont il évalue les résultats. Si les patients, les soignants et les évaluateurs ne sont pas au courant du traitement réellement reçu, nous parlons de triple aveugle.

### Risque de biais d'observation

Si les personnes qui évaluent les résultats d'une recherche sont au courant du traitement réellement reçu, leurs préjugés, leurs attentes, leurs espoirs pourraient fausser leur interprétation d'un résultat, particulièrement si le critère de jugement est plus subjectif.

#### Exemples :

Dans une synthèse méthodique des RCTs en orthopédie (traumatologie), Poolman et coll. constatent que 90 % des évaluations ne sont pas en insu<sup>4</sup>, ce qui expose à un risque de biais d'observation.

Dans des études évaluant l'intérêt d'une intervention chirurgicale (versus absence), la qualité de vie est évaluée en connaissance du traitement reçu (chirurgie ou non) et les capacités physiques sont évaluées en insu (exemple cité dans le Cochrane Handbook<sup>1</sup>).

### Effet d'un biais d'observation

Le non respect de l'insu lors de cette évaluation peut-il modifier, biaiser, les résultats pour certains critères (mais pas pour d'autres) ? Si oui, avec quelle ampleur ?

Une sommation des résultats de 7 méta-analyses de méta-analyses (= études méta-épidémiologiques), regroupant les données de 234 méta-analyses originales par Savovic et coll. avait étayé ce biais d'observation pour des critères de jugement subjectifs<sup>5</sup>.

Hrobjartsson et coll.<sup>6</sup> ont aussi montré dans leur méta-analyse de 21 RCTs rapportant des résultats pour des critères subjectifs binaires, résultats évalués à la fois en insu et non en insu, que l'évaluation non en insu exagérait les résultats, exagération exprimée en Odds Ratio, de 36 %.

Ce même groupe d'auteurs publie le même type de recherche pour des critères subjectifs évalués sur des échelles de mesure (scores) et non plus des critères

binaires<sup>7</sup>. Après une recherche exhaustive dans la littérature, avec contacts avec les auteurs originaux si nécessaire et recoupement des données dans les dossiers de la FDA, ils ont inclus 16 études (2 854 patients) évaluant le type de critères recherchés en insu ou non. Ils ont exclu les études sans distinction claire entre groupe expérimental et groupe contrôle, les études où les 2 types d'évaluation du même critère n'étaient faits que pour un sous-groupe de patients (sauf si sélection aléatoire), les études dans lesquelles les évaluateurs avaient accès aux résultats de l'autre type d'évaluation, les études dans lesquelles les évaluateurs étaient successivement en insu puis non en insu. Pour chaque étude, ils ont choisi une seule échelle de mesure, de préférence pour le critère primaire. Ils ont défini a priori 5 facteurs confondants possibles : un délai trop long entre les 2 types d'évaluation (en insu et non en insu), des évaluateurs de status différents (infirmières versus médecins par ex.), des procédures d'évaluation différentes (visuelle directe versus photographie par ex.), un risque substantiel de non respect de l'insu et une non similitude des patients évalués par les 2 méthodes. Ils ont sommé les différences en ampleur d'effet calculées dans chaque étude entre les 2 groupes, sommation en modèle d'effets aléatoires et avec pondération avec l'inverse de la variance. Dans 10 études (63 %) l'estimation des résultats est plus optimiste quand elle est faite non en insu. La différence d'ampleur d'effet pour l'ensemble des études entre les 2 types d'évaluation (insu ou non-insu) est de - 0,23 avec IC à 95 % de - 0,40 à - 0,06,  $p = 0,02$  et  $I^2$  à 46,3 % (non concordance modérée), ce qui correspond à une exagération de l'ampleur d'effet en moyenne de 68 % mais avec un IC à 95 % de 14 % à 230 %. En analyse de méta-régression univariée, les auteurs n'observent pas d'association statistiquement significative avec différentes caractéristiques d'études prédéfinies : haut score au point de vue subjectivité du critère de jugement, degré d'implication des évaluateurs non en insu dans l'étude, vulnérabilité du critère liée au non insu du patient.

Il faut souligner la distinction à respecter entre agrément inter-observateurs et biais d'observation. Dans l'étude reprise dans cette méta-analyse et qui présente la plus forte différence d'ampleur d'effet entre les deux groupes (- 1,10 avec IC à 95 % de - 1,54 à - 0,66, ce qui correspond à une exagération d'effet de 230 %), le coefficient de corrélation entre observateurs est de 0,87 (1 = corrélation parfaite) : il existe une excellente corrélation entre les différents observateurs mais le biais d'observation lié au non insu est très important.

### Conclusion

Cette recherche, de méthodologie complexe mais par ailleurs excellente, montre et chiffre l'importance possible du risque de biais lié à une évaluation non en insu de critères subjectifs dans les RCTs.

Références voir site web