GRILLE DE LECTURE

(Audet N. Ledère H., Pédagogie Médicale, 2001, 2, 206-212)

1. HABILETES REQUISES POUR LIRE DE FACON CRITIQUE : L'INVENTAIRE

A) DETERMINER LA VALEUR SCIENTIFIQUE D'UN ARTICLE

- **pertinence de l'article** par rapport à son questionnement
- évaluer la **justesse des données**

  . vérifier l'authenticité des faits
  . s'assurer de l'exhaustivité des données
  . distinguer un fait d'une opinion
  . évaluer si la taille de l'échantillon permet la généralisation des résultats
  . détecter les biais (erreurs systématiques)
  . dépister les relations de cause à effet erronées
  . reconnaître les affirmations non valides

- évaluer l'**interprétation des données**
  . critiquer le choix des tests statistiques
  . déterminer le niveau de preuve proposé par le plan de recherche
  . critiquer le choix du plan de recherche

- identifier les **généralisations fautives**

B) APRES LA LECTURE, REALISER UNE SYNTHESE portant sur l'utilité clinique des résultats

- en général
- pour sa propre pratique
- capacité à induire un changement de comportement professionnel
- utilité en tant qu'outil d'autoformation en cours de lecture (métacognition)

2. LA TECHNIQUE DE LECTURE D'UN ARTICLE : cadre de référence

CHOISIR CE QU'ON LIT

motivation : préoccupation fondamentale ou clinique

→ posséder une certaine maîtrise du sujet

→ disposer d'une stratégie efficace de recherche d'articles

- banques de données, Internet
- bibliothèque

- expert
- archivage d'articles

DECIDER DE LA LECTURE

- être concerné par le sujet
- être familier avec le contenu médical et les concepts de biostatistique et d'épidémiologie clinique utilisés.

IDENTIFIER LA PERTINENCE

a. **Comprendre les concepts méthodologiques utilisés**

b. **Evaluer la justesse des données**

- distinguer un fait d'un commentaire (IMRAD – introduction, méthodes, résultats, analyse, discussion)
- détecter les biais
- vérifier l'exhaustivité des données
- s'assurer de la qualité et de la pertinence des tableaux et figures

c. **Evaluer l'interprétation des données**

- niveau de preuve
- plan de recherche
- tests statistiques utilisés
- généralisation des résultats en accord avec les limites permises par la méthodologie par la population de référence

Cette étude est ou n'est pas valide sur le plan scientifique.

Cette étude est ou n'est pas utile pour la pratique médicale.

3. LE TITRE

Annonce le contenu de l'article avec le maximum de précision et de concision.

Clair, spécifique, non provocateur, il est autonome par rapport au reste de l'article tout en étant son reflet exact.

Court (10 à 15 Mots maxi) ; d'autant plus long que le sujet abordé est précis.

Attractif pour le lecteur

Sélectif pour la recherche bibliographique (garantie d'indexation)

N'y mettre que des mots informatifs

Construire un titre en utilisant au mieux **les mots clés de l'Index Médicus**

4 . MATERIEL ET METHODES

Suffisamment précis pour qu'un lecteur puisse reproduire ou vérifier le travail.
Le nœud de la rigueur scientifique et de la crédibilité.
Ne comporte ni résultats ni commentaire.

1er OBJECTIF

Le matériel de l'étude

- contient tous les détails nécessaires à l'interprétation des résultats
- doit permettre d'identifier les biais éventuels

La description comporte

- les critères d'inclusion ou d'exclusion
- série consécutive ou non
- travail prospectif ou rétrospectif, randomisé ou non, ouvert ou en insu, primo traitement ou non
- les éléments ayant permis la construction d'un échantillon ou d'une population
- caractéristiques précis de l'échantillon : âge, sexe, CSP, origine ethnique, symptomatologie, données morphologiques, groupes et sous-groupes
- accord du CCPPRB

2ème OBJECTIF

Ce que l'on cherche à évaluer

action d'un médicament, résultats d'une intervention, valeur diagnostique d'un examen

3ème OBJECTIF

Critères de jugement pour apprécier les résultats

- valeurs normales retenues
- tests statistiques
- précision : chiffrer un amaigrissement, quantifier une diarrhée
- indiquer le nombre de malades exclus de l'analyse (en donnant la raison) ou perdus de vue

En corollaire le chapitre « résultats » ne doit contenir que les résultats se référant directement aux critères de jugement.

5. DISCUSSION

Trois éléments majeurs :

a. Précise si le but du travail exposé à la fin de l'introduction a été atteint ou non (il est d'usage de résumer les principaux résultats)

b. Juge de la qualité et de la validité des résultats : la discussion critique est objective porte sur chacun des chapitres de l'article. Commentaire sur les biais.

c. Compare les résultats observés à ceux d'autres auteurs en cherchant à expliquer les différences et en soulignant l'apport original. Esquiver la critique personnelle.

Les erreurs

- **Répétitions** de ce qui a été dit dans l'introduction
- L' **inexactitude des citations**
- L'**usage des temps** : au passé tout ce qui relate un fait (résultat)
  au présent ce qui est opinion d'auteur et pour les notions bien établies

6. LE RESUME

(Rédaction Médicale Huguier M., Maisonneuve H., 1 vol., 3ème ed, 1998, DOIN)

Informatif, il doit présenter la substance des informations d'un article.

Il **répond aux 4 questions fondamentales**

- **pourquoi ce travail a été fait ?**
- **comment ?**
- **qu'est ce qui a été trouvé ?**
- **quelles conclusions ou généralisation ?**

- sa construction reprend la structure IMRAD : introduction, matériel et méthode, résultats, discussion (limité habituellement à une phrase de conclusion).
- **la première phrase** exprime l'idée directrice (sans répéter le titre)
- la **seconde** énonce la méthode de travail
- **utiliser des temps** passés sauf pour la conclusion et les hypothèses où le présent est de mise.
- **longueur** : moins de 250 mots (système de Vancouver)
- **les erreurs**
  . le résumé indicatif (au lieu d'informatif), réservé à la revue générale
  . éviter les abréviations
  . informations trop peu détaillées pour comprendre l'article

Autre structure possible (ou souhaitable)

a. **But de l'étude** : énoncer clairement l'hypothèse testée
b. **Protocole, méthode de l'étude** : thérapeutique, tests diagnostiques, étude de pronostic, évaluation économique, description de série de cas, enquête
c. **Lieu de l'étude** : ambulatoire, hôpital, institution
d. **Sélection des patients** : volontaires, malades, consécutifs, critères d'éligibilité
e. **Méthodes thérapeutiques** : durée, méthode, posologie
f. **Critère de jugement principal** : méthode d'analyse choisie avant la collecte des données
g. **Résultats** : avec intervalles de confiance et signification statistique
h. **Conclusion** : éviter les spéculations et les généralisations hâtives : s'en tenir aux faits

7. INTRODUCTION

Doit permettre de **comprendre pourquoi le travail a été entrepris** et quel était son **objectif**.

2 parties :

a) **Objectif informatif** : apporter au lecteur les éléments nécessaires et suffisants à la compréhension du travail (passerelle entre les connaissances de l'auteur et celles du lecteur)
b) **préciser le but du travail**

8. REFERENCES

- **Justifiant tout fait énoncé** : principe fondamental de la démarche scientifique
- **Sélectionner les références pertinentes**
- Elles **soutiennent les faits ayant amené les auteurs à formuler l'objectif du travail** (intro), renvoient aux méthodes déjà décrites (mat et méthodes) et augmentent les résultats critiques dans la discussion.
- **Pas de référence dans les résultats ou le résumé**.

Références à éviter : thèses, résumés de congrès, « communications personnelles », Articles « sous presse » ou « soumis à publication »

9. RESULTATS

Tous les résultats et rien que les résultats.

Même les résultats négatifs dans la mesure où ils apportent une information utile.

LES ERREURS

- donner les résultats qui ne sont pas cohérents avec le but du travail (confusion)
- faire des commentaires.

LES TABLEAUX ET FIGURES

Fournissent le **maximum d'information** sous une forme synthétique et claire

Evitent les répétitions entre l'exposé des résultats et leur rappel comme base de discussion dans le chapitre « discussion ».

Doivent **être informatifs** par eux-mêmes (légendes, titres...)

Doivent **être appelés dans le texte**.

Agence Nationale d'accréditation et d'évaluation en santé (ANAES) :

GUIDE D'ANALYSE DE LA LITTERATURE ET GRADATION DES RECOMMANDATIONS

Niveau de preuve des études	Grade des recommandations
Niveau 1 - essais comparatifs randomisés de forte puissance - méta-analyse d'essais comparatifs randomisés - analyse de décision basée sur des études bien menées	**A** **Preuve scientifique établie**
Niveau 2 - essais comparatifs randomisés de faible puissance - études comparatives non randomisées bien menées - études de cohortes	**B** **Présomption scientifique**
Niveau 3 - études cas-témoin **Niveau 4** - études comparatives comportant des biais importants - études rétrospectives - séries de cas - études épidémiologiques descriptives (transversale, longitudinale)	**C** **Faible niveau de preuve scientifique**

Les grands types de recherche

La plupart des études s'inscrivent dans l'une ou l'autre des catégories suivantes :

- *Thérapeutique* : mesurer l'efficacité d'un médicament, d'une intervention chirurgicale, d'une nouvelle méthode d'éducation des patients ou de toute autre procédure. Le protocole de choix est l'essai comparatif randomisé

- *Diagnostic* : montrer si un nouveau test diagnostique est valide (peut-on lui faire confiance ?) et fiable (fournit-il les mêmes résultats lorsqu'on le répète ?). Le protocole de choix est l'enquête transversale (*cross sectional survey*) où sont mis en œuvre simultanément le nouveau test et l'examen de référence (*gold standard*)
- *Dépistage* : estimer la valeur d'un test qui peut être appliqué à une vaste population et qui détecte une maladie à un stade présymptomatique. La méthode de choix est l'enquête transversale (*cross sectional survey*)

- *Pronostic* : prévoir comment évoluera une maladie diagnostiquée chez un individu à un stade précoce. La méthodologie de choix est l'étude de cohortes longitudinale
- *Etiologie* : démontrer si un agent présumé pathogène, une pollution environnementale par exemple, est effectivement impliqué dans la survenue d'une maladie. Le protocole de choix est soit l'étude de cohortes, soit l'étude cas-témoins, selon la fréquence de la maladie , mais les séries de cas peuvent également apporter des informations cruciales.

Grille d'évaluation de la qualité générale d'un article

1. Dans quel but l'étude a-t-elle été réalisée (à quelle question clinique est-elle censée répondre) ?
2. Quel est le type d'étude réalisé ?
   - étude primaire : expérience, essai comparatif randomisé, autre essai clinique comparatif, étude de cohortes, étude cas-témoins, enquête transversale, enquête longitudinale, description d'un cas ou d'une série de cas ?
   - étude secondaire : revue de la littérature, synthèse méthodique, méta-analyse, analyse décisionnelle, recommandation de pratique clinique, analyse économique ?
3. Le protocole de l'étude est-il approprié au domaine général de recherche concerné : traitement, diagnostic, dépistage, pronostic, étiologie ?
4. L'étude est-elle éthique ?

Grille d'évaluation de la qualité méthodologique d'un article

1. L'étude est-elle originale ?
2. Qui sont les sujets étudiés ?
   - Comment ont-ils été recrutés ?
   - Qui a été inclus et exclu de l'étude ?
   - Les sujets ont-ils été observés dans leurs conditions de vie habituelles ?
3. Le protocole de l'étude est-il judicieux ?
   - Quelle est l'intervention ou la stratégie évaluée ?
   - Quel(s) critère(s) est (sont)-il(s) choisi(s) pour en mesurer les effets ?
4. Les conditions expérimentales sont-elles adéquates ?
   - S'il s'agit d'un essai randomisé, la randomisation a-t-elle été vraiment aléatoire ?
   - S'il s'agit d'une étude de cohortes, d'une étude cas-témoins ou d'un autre type d'étude comparative non randomisée, le groupe témoin a-t-il été constitué de manière appropriée ?
   - Les groupes sont-ils comparables en tous points à l'exception de la variable étudiée ?
   - L'évaluation des résultats (ou la définition des « cas » dans une étude cas-témoins) a-t-elle été faite en « aveugle » ?
5. L'étude dispose-t-elle d'un effectif suffisamment important, d'une durée suffisamment longue et d'un taux de suivi suffisamment complet pour asseoir la crédibilité des résultats ?

Grille d'évaluation de la qualité des analyses statistiques d'un article

1. Les auteurs ont-ils planté le décor correctement ?
   - Ont-ils vérifié la comparabilité des groupes et les ont-ils ajustés pour les différences initiales éventuellement présentes ?
   - Quel type de données ont-ils recueilli, et les ont-ils analysées par des tests statistiques appropriés ?
   - Si les tests statistiques utilisés ne sont pas courants, pourquoi les auteurs y ont-ils fait appel ?
   - Les données ont-elles été analysées conformément au protocole initial ?
2. Données appariées, latéralité des tests et valeurs atypiques
   - Les données appariées ont-elles été analysées par les tests pour données appariées ?
   - Un test bilatéral a-t-il été effectué chaque fois qu'une intervention pouvait susciter un effet positif ou négatif ?
   - Les valeurs atypiques ont-elles été examinées sous l'angle du bon sens et traitées par les ajustements statistiques appropriés ?
3. Corrélation, régression et causalité
   - La corrélation a-t-elle été distinguée de la régression ? Le coefficient de corrélation r a-t-il été calculé et interprété correctement ?
   - Des hypothèses ont-elles été formulées sur la nature et le sens du lien de causalité ?
4. Probabilités et intervalles de confiance
   - Les valeurs de p ont-elles été calculées et interprétées de façon appropriée ?
   - Les intervalles de confiance ont-ils été calculés et ont-ils été pris en compte par les conclusions des auteurs ?
5. Les auteurs ont-ils exprimé les résultats en termes de bénéfices ou de nuisances prévisibles pour les patients, autrement dit sous forme de :
   - réduction relative du risque ;
   - réduction absolue du risque ;
   - nombre de sujets à traiter ;
   - rapport des cotes.

TABLEAU 4-I. – GRILLE DE LECTURE

L'INFORMATION EXISTE-T-ELLE POUR CHACUNE DES 8 QUESTIONS ?	LA FAÇON D'ABORDER LA QUESTION EST-ELLE CORRECTE ?	SI NON, CELA MENACE-T-IL LA VALIDITÉ DE L'ÉTUDE ?
1. Objectif – pronostic-évolution – test diagnostique – impact d'une intervention – étiologie-causalité	– Y a-t-il une hypothèse ?	
2. Types d'étude – rapport de cas-série de cas – étude transversale – étude cas-témoins – étude de cohorte – essai contrôlé	Le type de l'étude est-il approprié à la question posée ?	Si non, les résultats de l'étude sont-ils totalement inutiles ?
3. Facteurs étudiés – exposition – intervention – test diagnostique	Comment sont-ils mesurés ? – Même méthode de mesure chez tous les sujets ? dans tous les groupes ? – Méthode à l'aveugle ? Y a-t-il une comparaison indépendante avec le gold standard?	– Si non, ce biais de mesure menace-t-il la validité de l'étude ? – Idem. Si non, ce biais menace-t-il la validité de l'étude ?
4. Critères de jugement	Comment sont-ils mesurés ? – Idem question 3. Tous les critères de jugement pertinents sont-ils évalués ?	– Idem question 3. Si non, ceux qui ont été oubliés sont-ils importants ?
5. Population source et sujets étudiés	– La sélection est-elle correcte ? – Y a-t-il randomisation ? – Les groupes diffèrent-ils par des caractéristiques autres que les facteurs étudiés ? – Quelle est la proportion de sujets atteignant la fin du suivi ?	– Si non, ce biais menace-t-il la validité externe? – Si non, ce biais menace-t-il la validité interne ? – Si elle n'est pas optimale, la validité interne est-elle menacée ?
6. Facteurs de confusion potentiels	– Sont-ils tous envisagés ? – Sont-ils bien contrôlés ?	– Si non, cela invalide-t-il l'étude ?
7. Analyses statistiques – Intervalle de confiance ? – Test statistique ? si résultats positifs si résultats négatifs – Force de l'association	– Taille de l'échantillon suffisante ? – Cliniquement intéressant ? – Puissance du test ? Taille de l'échantillon ?	– Si non, les résultats sont-ils inutiles ? – Si non, l'étude est-elle utile ? – Si insuffisant, l'étude est-elle utile ou non concluante ?
8. Conclusions des auteurs ? – Réponses aux questions ? – Vérification de l'hypothèse ? – Objectif atteint ?	Les conclusions répondent-elles à l'objectif ?	*En somme :* – Les résultats sont-ils acceptables appliqués à la population source ? **= validité** – Les résultats sont-ils acceptables pour votre propre pratique ? **= applicabilité**

L'information existe-t-elle pour chacune de ces 8 questions ?	La façon d'aborder la question est-elle correcte ?	Impact sur les conclusions
1.Objectif - Pronostic – évolution - Test diagnostique - Impact d'une intervention - Etiologie – causalité	- Y a-t-il une hypothèse ? - L'objectif principal est-il clairement défini ?	
2. Type d'étude - Essai contrôlé randomisé - Etude de cohorte - Etude cas-témoins - Etude transversale - Rapport de cas – série de cas	- Le type de l'étude est-il approprié à la question posée ? - S'agit-il du type d'étude apportant le niveau de preuve le plus élevé	- Si non, les résultats de l'étude sont-ils totalement inutiles ?
3. Facteur(s) étudié(s) - Exposition - Intervention - Test diagnostique	Sont-ils bien décrits ? Comment sont-ils mesurés ? - Même méthode de mesure chez tous les sujets ? dans tous les groupes ? - Méthode à l'aveugle ? Y a-t-il une comparaison indépendante avec l'étalon ?	Sinon ce biais de mesure menace-t-il la validité de l'étude ? *Idem* Sinon ce biais menace-t-il la validité de l'étude ?
4. Critère(s) de jugement	Le critère de jugement est-il unique ? Si non, a-t-on défini un critère principal Comment sont-ils mesurés ? - *Idem* question 3 Tous les critères de jugement pertinents sont-ils évalués ? En cas de critères multiples, les résultats sont-ils convergents ?	- *Idem* question 3 Si non, ceux qui ont été oubliés sont-ils importants ?

5. Population source et sujets étudiés	- Les critères d'inclusion et d'exclusion sont-ils précisés et pertinents ? - Y a-t-il des perdus de vue ? Si oui, leur nombre et leur répartition sont-ils précisés ? - Y a-t-il des non-réponses ? Comment ont-elles été prises en compte ? - A-t-on vérifié la robustesse des conclusions selon les hypothèses faites pour les non-réponses ? - Y a-t-il randomisation ? A-t-elle été effectuée correctement ? Les effectifs permettant de vérifier les caractéristiques des sujets inclus et leur bonne répartition sont-ils présentés ? - Les groupes diffèrent-ils par des caractéristiques autres que les facteurs étudiés ? - Quelle est la proportion de sujets atteignant la fin du suivi ?	- Si des biais sont à craindre, pour chacune des questions précédentes, cela menace-t-il : - la validité interne ? - la validité externe ?
6. Facteurs de confusion potentiels et biais	- Sont-ils tous envisagés ? - Sont-ils bien contrôlés ? - Biais d'information (de mesure, cf 3.) ? Erreurs différentielles ? Erreurs non différentielles ? - Biais de sélection ?	- Si non, cela invalide-t-il l'étude ? - Situation rendant l'étude inexploitable, si importantes - Conclusion possible uniquement si résultat significatif - L'interprétation est-elle encore possible ?

7.Résultats Type d'analyse utilisée ? Intervalles de confiance ?	Est-elle adaptée ? A-t-on privilégié la présentation des estimations accompagnées d'un intervalle de confiance ?	- Si non, peut-on les calculer à partir des données présentées ?
Tests statistiques	- Réalisés uniquement pour les hypothèses indiquées au départ ? - Respect des conditions d'application ?	- Le risque global d'erreur de première espèce est-il conservé ?
- Si résultats positifs - Si résultats négatifs	- Cliniquement intéressant ? - Taille de l'échantillon suffisante (puissance) ?	- Si non, l'étude est-elle utile ? - Si insuffisant, l'étude est-elle utile ou non concluante ?
Force de l'association	- Est-elle indiquée ? - L'indicateur choisi est-il pertinent ? - Peut-on vérifier les résultats à partir des données présentées ?	
Cohérence externe	Les résultats sont-ils cohérents - avec ceux obtenus par d'autres études épidémiologiques ? - avec ceux obtenus dans d'autres disciplines, par d'autres méthodes ?	
8. Conclusions des auteurs ? Réponses aux questions ? Vérification de l'hypothèse ? Objectif atteint ?	- Les conclusions répondent-elles à l'objectif ?	Les résultats sont-ils acceptables appliqués à la population source (validité) ? Les résultats peuvent-ils être extrapolés aux situations qui vous intéressent pour votre propre pratique (applicabilité) ?

GRILLE DE LECTURE D'UN ESSAI CLINIQUE

d'après l'enseignement du DIU de Formation des Investigateurs de l'Essai Clinique (DIU-FIEC)

La lecture critique d'une publication d'essai clinique procède en trois temps :

- identification de l'objectif de l'étude,
- analyse méthodologique,
- pertinence des résultats pour la pratique clinique.

1. OBJECTIF

L'essai ne peut répondre qu'à une seule question : cet **objectif (principal)** doit être clairement décrit. Sa définition fera choisir un **critère de jugement principal** qui sera, entre autre, utilisé pour calculer le **nombre de sujets nécessaire**. Des **objectifs secondaires** peuvent y être associés : ainsi une analyse de la sécurité d'emploi du médicament (la tolérance) peut s'ajouter à un objectif principal ciblé sur l'évaluation de l'efficacité.

Il est souhaitable que l'utilité thérapeutique de la question posée soit explicitée, ou du moins située dans le contexte actuel (ces données étant référencées dans la bibliographie).

Schématiquement, il faut vérifier que :

- le but de l'essai est défini,
- de façon compréhensible ;
- la réponse à la question posée pourrait avoir une application clinique.

2. METHODOLOGIE

Un essai **prospectif**, **contrôlé** (vs placebo ou vs un **traitement de référence** reconnu), **randomisé** selon un **plan expérimental** adéquat mené en **double-insu** sur un **échantillon** parfaitement défini dont l'effectif aura été **calculé** selon la **puissance** choisie, utilisant un **critère de jugement** adapté à l'objectif, offrira **un niveau de preuve** optimal.

Ces paramètres seront considérés selon 3 questions (a, b, c)

a) Comment les réponses aux traitements testées vont-elles être mesurées ?

- Selon un plan expérimental correctement choisi (groupes parallèles, chassé-croisé, factoriel, séquentiel)

- Le critère d'évaluation (principal) répond bien à l'objectif de l'étude. Le choix des horaires de sa mesure, la sensibilité et la spécificité de l'instrument de mesure, éventuellement la variabilité inter-individuelle (paramètre utile pour le calcul du nombre de sujets nécessaire) sont connus. Autrement dit, il faut vérifier que : un critère principal est défini pour mesurer l'efficacité ; ce critère principal a une signification médicale claire ; il est adapté au but de l'essai.
  Chaque fois que possible c'est un critère quantitatif qui sera choisi (plutôt que qualitatif).

- L'échantillon de la population étudiée est parfaitement défini par des **critères d'inclusion et de non-inclusion**. Il faut se souvenir que des critères d'inclusion étroits permettront d'obtenir une homogénéité des patients inclus favorable à la détection d'une significativité statistique mais que les résultats obtenus sur cet échantillon très particulier pourront ne pas être extrapolables à la généralité des patients souffrant de l'affection considérée.

- Le choix du **traitement de référence** (ou traitement « contrôle ») répond aux données actuelles du traitement de l'affection considérée, qu'il s'agisse de l'emploi d'**un placebo** ou d'une spécialité pharmaceutique qui devra alors être utilisée à la posologie et pour la durée reconnues dans l'indication faisant l'objet de l'étude.

b) Comment la comparabilité des groupes est-elle satisfaite ?

- La répartition des traitements est faite par **tirage au sort**. La méthode utilisée est correcte (le plus souvent : table de nombre au hasard) S'il existe une stratification, elle est justifiée sur le plan clinique et ne multiplie par les sous-groupes de façon excessive et préjudiciable à la puissance de l'étude.

- Le **double-insu** (éventuellement **simple insu**) est bien assuré, au besoin avec la participation d'observateurs externes (lorsque la mesure d'un critère de jugement risque de lever l'aveugle). La méthode du « double-placebo » est éventuellement utilisée, à bon escient.

- Le nombre de patients sortis de l'étude ou perdus de vue n'est pas différent selon les groupes (ou les périodes).

c) *Comment la puissance de l'essai est-elle assurée ?*

- Le **nombre de sujet nécessaire** a été calculé. Les risques α ($1^{ère}$ espèce) et β ($2^{ème}$ espèce) acceptés sont indiqués. Cet effectif est atteint.

- Le nombre de sorties d'étude (< 10 à 15 % des inclus), de **perdus de vue** (< quelques %), ou de valeurs manquantes, n'est pas invalidant.

- L'**observance** des traitements comparés à été mesurée. Les moyens de favoriser cette observance ont été mis en œuvre.

- En cas d'**essai multicentrique**, le choix des centres participants est raisonnable et des efforts ont été faits pour standardiser au niveau de chacun d'eux les procédures de l'étude. Un tirage au sort séparé, équilibré par centre, a été fait.

- Le recueil des données et leur saisie ont été organisés de façon à réduire les erreurs de transcription.

3. ANALYSE DES RESULTATS

a) La comparabilité des groupes expérimentaux à l'inclusion (au début de l'étude) est vérifiée. Le choix des critères de comparabilité est raisonnable.

b) Des tableaux rendant compte des effectifs de sujets inclus, sortis d'étude, perdus de vue, avec les effectifs utilisés pour une analyse en **intention de traiter** ou en « **per protocole** » sont fournis. Les chiffres qu'ils mentionnent sont concordants, en particulier avec le texte de la publication.

c) Le test statistique utilisé pour analyser la différence entre les valeurs du critère de jugement principal recuillies pour chaque groupe expérimental est bien choisi. L'analyse statistique porte sur la totalité des patients inclus dans l'essai (analyse en intention de traiter). Il existe une différence avec un $p < 0.05$. Ce résultat est retrouvé par l'analyse per protocole. Cette différence statistiquement significative est cliniquement intéressante.

d) Les résultats d'éventuelles analyses en sous-groupe ne sont donnés qu'à titre informatif.

e) Les résultats de l'analyse d'éventuels critères de jugement secondaires sont cohérents.

4. CONCLUSION

Elle reprend les faits démontrés sans donner d'opinion ou formuler de nouvelle(s) hypothèse(s). Elle indique si le résultat obtenu peut être extrapolé à l'ensemble des formes cliniques de l'affection traitée ou doit rester limité au type d'échantillon recruté pour l'étude. Elle souligne l'éventuel bénéfice thérapeutique pratique qui résulte de l'essai.

* Les notions explicitées dans le glossaire figurent en **caractères gras** lors de leur première citation dans cette grille.

LECTURES CONSEILLEES

- Dossier « Essais cliniques » paru dans la Revue du Praticien, 2000, 50, p. 827 à 869, et en particulier : BERGMANN J.F. et CHASSANY O. Lire un compte-rendu d'essai clinique, p. 838 à 845.
- BOUVENOT G., VILLANI P., AMBROSI P. Lecture critique de la publication d'un essai clinique. Presse Med., 2002, 31 : 1061-8.
- Dossier « La vie du médicament » paru dans la Revue du Praticien, 2002, 52 : 480-515

GRILLE DE LECTURE D'UN ARTICLE ORIGINAL EVALUANT UN TEST DIAGNOSTIQUE

B. Goichot, S. Rohr

NB : par « test », on entend toute information dont le résultat peut être utilisé dans une démarche décisionnelle, le plus souvent dans un but diagnostique. Ce peut être un examen biologique ou d'imagerie, un signe clinique, une procédure ou une stratégie (c'est à dire la combinaison de plusieurs tests)

1. Identifier ce qui est évalué : examen biologique ou autre, signe clinique, stratégie....
   - Le test évalué est-il suffisamment bien décrit pour être reproduit ?
2. Identifier l'examen de référence (« gold standard »)
   - Le test étudié doit être comparé à l 'examen le plus performant disponible pour affirmer le diagnostic. Le choix de cet examen de référence est capital, il doit être justifié dans le texte avec des références adéquates. Lorsque cet examen n'existe pas, cela peut être une vérification autopsique ou chirurgicale, plus rarement une évolution clinique qui confirme a posteriori le diagnostic.
   - Tous les patients étudiés doivent avoir les deux tests, celui qui est évalué et celui de référence, et l'interprétation de chacun de ces tests doit être faite de façon indépendante, en « aveugle ».
   - La maladie et l'absence de maladie sont elles bien définies ?
3. La population étudiée est elle correctement décrite ?
   - Quelle est la fréquence de la maladie dans cette population ? Cette fréquence sera elle identique dans les populations auxquelles les auteurs destinent ce test ?
   - S'il y a des stades de gravité différente pour cette maladie, sont-ils mesurés ?
4. La reproductibilité du test est-elle analysée (variabilité intra/inter observateur) ? L'interprétation du test dépend elle de l'expérience d'un examinateur (par ex. examen radiologique) ?
5. Comment sont exprimées les qualités diagnostiques du test : sensibilité, spécificité, valeurs prédictives, rapports de vraisemblances (voir glossaire) ?
6. Les intervalles de confiance sont-ils fournis ?
7. L' « applicabilité » du test est elle discutée en tenant compte : de la fréquence de la maladie dans d'autres populations ? de la faisabilité technique ? d'une évaluation économique ? Ce test a t-il été placé dans le contexte des autres examens utilisés et de leur séquence ?

RAPPELS

	Malade	Non malade	Total
Test positif	VP	FP	VP+FP
Test négatif	FN	VN	FN+VN
Total	VP+FN	FP+VN	n

Lecture verticale : Se = VP/(VP+FN) Sp = VN/(VN+FP)
Lecture horizontale : VPP = VP/(VP+FP) VPN = VN/(VN+FN)

Sensibilité (Se) = probabilité d'avoir un test positif quand on est malade
VPP = probabilité d'être malade quand on a un test positif

Spécificité = probabilité d'avoir un test négatif quand on n'est pas malade
VPN = probabilité de ne pas être malade quand on a un test négatif

Sensibilité et spécificité décrivent les qualités *intrinsèques* du test.
Les valeurs prédictives dépendent des qualités intrinsèques du test et de la *probabilité pré-test*, c'est à dire de la prévalence de la maladie dans la population étudiée. Elles décrivent l'utilité diagnostique du test dans *une population donnée.*

Les *rapports de vraisemblance* (LR likehood ratio) décrivent l'apport d'un test au diagnostic :

Le rapport de vraisemblance positif (L) d'un test est la vraisemblance de la présence de la maladie lorsque le test est positif. Il varie de 0 à l'infini. Si LR = 1, cela signifie que la probabilité après avoir fait le test est identique à la probabilité pré-test, donc que le test n'apporte rien au diagnostic. Plus il est élevé, plus le « gain diagnostique » est important.

Le rapport de vraisemblance négatif (λ) est la vraisemblance d'être malade lorsque le test est négatif. Il varie entre 0 et 1. Plus il est proche de 0, plus il permet d'exclure le diagnostic.

L = Se/(1-Sp) = (VP/malades)/(FP/non malades)
λ = (1-Se)/Sp = (FN/malades)/(VN/non malades)

Les rapports de vraisemblance sont indépendants de la prévalence de la maladie et sont un bon indice de la « valeur diagnostique » d'un test.

GLOSSAIRE DE TERMES SOUVENT UTILISES DANS LA REDACTION DES ARTICLES MEDICAUX

Pierre MEYER et Jean-Louis IMBS

Ce glossaire a pour objectif de rappeler le sens des principaux termes utilisés dans la rédaction d'article médicaux. Il va à l'essentiel des notions nécessaires dans la méthodologie biostatistique tout en limitant l'utilisation des formules et cite les principales composantes de l'essai clinique contrôlé.

Analyse en intention de traiter

En route rigueur, ce que l'essai contrôlé compare, c'est l'intention de traiter par un traitement A à l'intention de traiter avec un traitement B. Il faut donc conserver dans l'analyse tous les patients inclus dans l'étude selon leur répartition au hasard (randomisation). Cette allocation aléatoire est l'une des garanties que les deux groupes de patients (traité par A ou par B) sont comparables. Toute soustraction de l'analyse de patients (sortis d'étude, ou ne satisfaisant pas aux critères d'inclusion ou pour lesquels des valeurs sont manquantes, ou encore perdus de vue) risque d'induire un biais.

Cette approche prend la réalité clinique en compte : il est par exemple impossible de continuer à traiter un patient par un traitement qu'il ne supporte pas uniquement pour des exigences d'analyse statistique. En corollaire, il faut signaler que cette analyse ne permet que de mesurer l'efficacité clinique, c'est à dire une efficacité définie dans le cadre de la pratique courante. L'analyse en intention de traiter peut être associée avec une analyse selon le protocole (ou "per protocole") dans laquelle des patients peuvent être exclus de l'analyse pour des raisons variées comme celles citées ci-dessus.

Analyse per protocole

Cette analyse ne porte que sur les patients qui ont terminé et réalisé entièrement l'étude et pour lesquels toutes les mesures prévues sont disponibles.

Biais d'attrition

Le biais d'attrition intervient lorsque des patients randomisés sont exclus de l'étude ou écartés de l'analyse. Ce biais est fréquent et assez facile à détecter lorsque les effectifs sont suffisamment bien définis dans l'article. Dans la pratique ce biais est limité par une bonne organisation de l'étude.

Biais d'évaluation

Ce biais survient lorsque l'évaluation du résultats du traitement est influencée par des facteurs subjectifs l'étude n'étant pas en double insu (double aveugle). Il faut parfois faire intervenir un évaluateur externe pour mesurer objectivement le critère de jugement (encore appelé triple aveugle).

Biais de confusion

Le biais de confusion est le biais entraîné par l'absence de la prise en compte d'un facteur pouvant influencer le résultat, de ce fait ce facteur tout en étant correctement réparti peut entraîner une dispersion importante cachant une différence réelle entre groupes. Des méthodes d'ajustement peuvent prendre en compte un facteur de confusion, encore faut-il l'avoir identifié.

Biais de sélection

Ce biais existe lorsque les deux (ou plusieurs) groupe(s) construits ne sont pas comparables. C'est le cas lorsque l'on effectue une comparaison avec un groupe de référence (témoin) non construit à partir de la même population que celle des autres groupes. Deux points sont à souligner :

- La randomisation doit être effectuée selon les bonnes règles (p.ex. en centralisée si l'étude est multicentrique, équilibrée ou stratifiée si nécessaire).
- La bonne comparabilité doit être vérifiée sur certains facteurs car la randomisation ne donne pas une garantie absolue de comparabilité.

Biais de suivi

Ce biais est introduit lorsque la comparabilité n'est pas maintenue tout au long de l'étude. Les causes de ce biais sont nombreuses comme par exemple les déviations au protocole, les traitements concomitants qui peuvent être différents, les arrêts du traitement pour diverses raisons.

Caractérisation d'une grandeur qualitative

Une grandeur qualitative peut être à deux niveaux, comme la présence ou absence d'un signe (variable dichotomique), ou à plus de deux niveaux, comme l'évaluation de la gravité d'une maladie (variable polychotomique). La distribution d'une variable dichotomique suit une loi binomiale, celle d'une variable polychotomique suit une loi multinomiale. On peut caractériser de telles variables par la fréquence et l'écart type de chaque modalité. On peut également les caractériser par la fréquence et par l'intervalle de confiance de la fréquence.

Caractérisation d'une grandeur quantitative

Une grandeur quantitative est caractérisée par sa distribution (gaussienne ou autre), sa valeur moyenne et son écart type. D'autres paramètres sont parfois utilisés comme les valeurs extrêmes et les quantiles (quartiles, déciles). Une distribution différente d'une distribution gaussienne peut être transformée pour se ramener à une loi de Gauss et bénéficier des tests paramétriques. Lorsque aucune transformation n'est souhaitée, il convient d'utiliser les tests non paramétriques.

Causalité

Cette notion est employée lorsque l'on souhaite prouver que la variation d'un facteur explicatif est la cause des variations d'un facteur à expliquer. En recherche clinique, la seule méthode permettant de prouver une relation de cause à effet est la méthode expérimentale. Rechercher une relation de cause à effet entre deux variables nécessite la construction d'une étude expérimentale dans laquelle l'une des variables est contrôlée, la variable à expliquer étant simplement mesurée.

Remarque : L'existence d'une liaison entre deux variables obtenues autrement que par la méthode expérimentale ne prouve pas la causalité. Lorsqu'un coefficient de corrélation entre deux grandeurs quantitatives est statistiquement différent de zéro, on a uniquement montré que les deux variables n'étaient pas indépendantes (lorsque les distributions sont gaussiennes).

Comparabilité des groupes

Un des critères dans un essai contrôlé est la comparabilité des groupes. Les groupes doivent être obtenus par randomisation à partir d'une population initiale clairement définie. Il faut veiller à maintenir cette comparabilité afin d'éviter d'introduire des variations non

contrôlées dans l'étude ce qui conduirait à des biais de suivi, d'attrition ou d'évaluation. Il est important de vérifier, une fois l'étude terminée et ses résultats saisis, qu'il n'y avait effectivement pas de différence entre les groupes au début de l'étude. Exceptionnellement, le hasard (la randomisation) peut faire qu'une telle différence existe, rendant alors très difficile l'interprétation de l'étude.

Comparaison de plusieurs groupes

Dans un essai contrôlé on peut être amené à comparer la moyenne de plusieurs groupes. La méthode à utiliser consiste à construire une étude basée sur l'analyse de la variance à un facteur à plusieurs modalités (plusieurs groupes). L'analyse de la variance a pour but de comparer les moyennes tout en gardant le risque α à 5%. La conclusion de l'analyse est que l'une ou plusieurs moyennes sont statistiquement différentes. L'analyse est complétée par la localisation de la différence (test de Newman Keuls, test de Scheffe). La mauvaise méthode consisterait à comparer les groupes entre eux (comparaison multiple).

Comparaison multiple

En répétant les tests de comparaison sur les mêmes données (chaque groupe est utilisé plusieurs fois pour le test) on modifie le risque α initial fixé. Par exemple, si le risque α est fixé à 0,05 pour un test, pour deux tests ce risque sera de 0,08 et pour trois tests il sera de 0,11. Une façon de corriger cette modification du risque est de protéger le test selon la méthode de Bonferoni en baissant le risque α initial en fonction du nombre de tests prévus (prendre un risque $\alpha' = \alpha / n$ pour n tests). La bonne méthode consiste à effectuer une analyse de la variance.

Consentement éclairé

Libre acceptation (exprimée par écrit) d'une personne envisageant de participer à un essai donné. Cette acceptation formelle ne doit être demandée qu'après information (dite "éclairée") de la personne, précisant les objectifs, les bénéfices, les risques et les inconvénients potentiels, liés à l'essai ; la personne doit également être informée de ses droits et responsabilités. Son acceptation doit être recueillie avant son inclusion dans l'étude.

Critère de jugement (voir aussi critères d'efficacité)

L'essai contrôlé a pour objectif de décider si deux ou plusieurs groupes sont statistiquement différents pour une variable définie. Cette variable constitue le critère principal de jugement. Celui-ci peut-être qualitatif ou quantitatif. Dans un essai contrôlé, il n'y a qu'un seul critère de jugement principal : c'est par rapport à lui que les effectifs nécessaires sont définis. D'autres paramètres que cette variable principale peuvent être intéressants à étudier : ils constituent des variables secondaires et aucun pari n'est fait sur elles.

Critères d'efficacité (en anglais : end points)

Le ou les critères permettant de mesurer les effets d'un médicament dépendent des buts de l'essai et, évidemment, de la classe du produit étudié. Un critère de jugement principal de l'efficacité doit être choisi de façon à répondre à l'objectif principal de l'étude. En somme, il s'agit du mode de mesure du bénéfice que le patient peut retirer du médicament.

Il existe trois types de critères d'efficacité : clinique, intermédiaire, de substitution.

Critères cliniques

Ils traduisent une amélioration quantitative ou qualitative de sa santé perçue par le patient, tels que :

- amélioration de la probabilité de survie ;

- prévention d'un état morbide ;
- disparition d'un symptôme douloureux, gênant ou invalidant ;
- guérison plus rapide ;
- amélioration de la qualité de la vie.

Critère intermédiaire

Amélioration du résultat d'un examen complémentaire paraclinique (radiologique par ex.) ou biologique (glycémie par ex.)

Critère de substitution (en anglais : surrogate end point)

Certains critères intermédiaires peuvent se substituer au critère clinique si :

- la preuve est faite que l'amélioration du critère de substitution coïncide certainement avec l'amélioration du critère clinique ;
- le critère de substitution est d'acquisition plus facile que celle du critère clinique.

Un exemple en est la baisse de la pression artérielle dans le traitement de la maladie hypertensive).

Description d'une population

La description de la population étudiée permet au lecteur d'apprécier le contexte dans lequel l'étude est effectuée. Elle est nécessaire afin de connaître avec précision les critères d'inclusion et d'exclusion de l'étude, le lecteur pourra ainsi évaluer la portée de l'étude. Une description complète facilitera la comparaison des résultats avec d'autres études et pourra faciliter une méta analyse.
La population étant définie par le problème étudié, l'étude sera en général effectuée sur un échantillon représentatif de cette population.

Double – insu (ou double aveugle, traduit littéralement de l'anglais "double-blind")

Cette notion s'applique à l'essai contrôlé organisé de manière que ni le patient, ni le médecin, ni le personnel soignant, ni l'entourage du malade ne connaissent celui des traitements testés que reçoit un patient donné. En revanche, les médicaments ou produits testés dans l'essai considéré sont bien sûr connus de tous, y compris des patients qui ont été "éclairés" par une information appropriée.
L'essai contrôlé en double insu a pour objet de limiter les biais qui pourraient s'introduirent dans l'étude par la connaissance du traitement car patients et médecins pourraient alors manquer d'objectivité.

Ecart-type

Cette notion caractérise la variabilité d'une grandeur quantitative. Cette grandeur est utilisée dans de nombreux calculs en statistique, par exemple pour calculer un intervalle de confiance, définir un paramètre discriminant pour effectuer un test, calculer des effectifs nécessaires, etc…

Echantillon

L'objectif de la recherche clinique est d'obtenir des résultats (des connaissances) présentant un caractère général (universel) afin de pouvoir utiliser cette connaissance sur l'ensemble d'une population. L'étude qui amènera cette connaissance ne peut que très rarement être effectuée sur toute la population, c'est pour cette raison qu'elle sera effectuée sur un échantillon représentatif de la population. Constituer un échantillon c'est tirer au hasard une sous-population d'effectif n dans une population (population parente) de taille N.
Une taille suffisante de l'échantillon augmente la précision des paramètres estimés.

C'est à partir des connaissances obtenues sur l'échantillon que l'on induit les connaissances sur la population. On détermine certaines grandeurs à partir de l'échantillon, celles-ci constituent des estimations des grandeurs correspondantes de la population. Lorsque l'échantillon est bien constitué, la statistique inductive nous donne des indications sur la population avec un certain risque d'erreur. On peut ainsi, à partir de l'échantillon, caractériser la population en effectuant une estimation ponctuelle de certaines grandeurs ou une estimation par intervalle de confiance de ces même grandeurs. Cette démarche nécessite au préalable la vérification de certaines hypothèses.

Remarque : Tirer au hasard ou au sort, ou randomiser (de "random", hasard en anglais), signifie éliminer toute subjectivité (biais) dans le choix de l'échantillon. Dans la pratique, pour obtenir un échantillon correct d'une population définie par des critères d'inclusion et d'exclusion, il faut se servir de tables numériques fournissant des séries de nombres aléatoires.

Effectifs nécessaires : voir "nombre de sujets nécessaire"

Essai contrôlé

Type d'étude demandant une organisation bien codifiée dans laquelle les échantillons retenues pour l'étude proviennent d'une population commune par randomisation. Certaines grandeurs sont contrôlées par la définition de la population, les autres grandeurs étant réparties de façon aléatoire dans les groupes. Dans un tel essai, on peut mettre en évidence des relations de cause à effet. Cette étude est démonstrative, la différence mise en évidence provient du facteur étudié.
Selon l'organisation, l'essai peut être en simple aveugle (simple insu) ou en double aveugle (double insu) selon que le patient ou le patient et le médecin ignorent quel est le médicament attribué.

Etude cas – témoins ; Enquête cas - témoins

Une telle étude est classée dans les enquêtes. Une enquête n'est jamais démonstrative. Dans une enquête de ce type, on connaît les cas (les malades sont souvent ceux de nos hôpitaux) et l'on cherche des témoins (non malades) ressemblants le plus possible aux malades. On dispose alors de deux groupes, plus ou moins comparables, sur lesquels on recherche le facteur à étudier, par exemple l'exposition à un risque. Ce type d'enquête permet de construire une table à quatre cellules contenant les effectifs pour les situations M E, $\bar{M}$ E, M $\bar{E}$, $\bar{M}$ E. On peut calculer s'il existe une liaison entre le facteur maladie (M) et le facteur exposition (E). Il est impossible de calculer le risque relatif, on peut cependant obtenir une grandeur appelée Odds Ratio (rapport des cotes) qui converge vers le risque relatif lorsque la maladie est rare. Ce type d'enquête est plus facile à réaliser que l'enquête sur échantillon représentatif. Il est l'objet cependant d'un certain nombre de biais.
Ce type d'étude est fréquemment effectué chez les patients hospitaliers, certains patients malades étant pris comme " cas ", d'autre patients non malades étant pris comme " témoins ". Cette situation pose souvent de nombreux problèmes d'interprétation.

	E	$\bar{E}$
M	a	b
$\bar{M}$	c	d

Les termes a, b, c, d représentent des effectifs. L'effectif total est (a + b + c + d).

Il faut remarquer que les taux de malades et de non malades ne sont pas représentatif de la population par conséquent, les rapports [a / (a + c)] et [b / (b + d)] n'ont pas de sens (double trait horizontal rappelant cette notion, pas de calcul du risque relatif mais de l'odds ratio).

Etude d'une série de cas

Ce type d'étude correspond à l'analyse d'une série de cas sans pour autant avoir un objectif de comparaison avec un groupe témoin. L'analyse d'une telle série relève essentiellement de la statistique descriptive. Elle permet de faire le point sur une méthode dans un but descriptif ou dans un objectif préliminaire de faire un étude plus structurée dans un deuxième temps. On accorde généralement un faible niveau de preuve à ce type d'étude (4 sur une échelle allant de 1 à 4 (critères de l'ANAES)).

Etude de cohorte ; Enquête de cohorte

Une telle étude est classée dans les enquêtes. Une enquête n'est jamais démonstrative. Ce type d'enquête encore appelé "enquête exposés - non exposés" dans laquelle on suit des individus dans le temps, les uns étant exposés à un risque et les autres non. On observe l'arrivée des maladies et au bout d'un certain temps d'observation, on effectue l'analyse. Ce type d'enquête permet de construire une table à quatre cellules contenant les effectifs pour les situations M E, $\bar{M}$ E, M $\bar{E}$, $\bar{M}$ E. Il est possible de déterminer s'il existe une liaison statistique entre les deux facteurs (calcul du Chi²) et l'on peut calculer le risque relatif. La liaison obtenue n'est cependant pas une liaison de cause à effet. Ce type d'enquête est plus facile à réaliser que l'enquête sur échantillon représentatif, il demande cependant une longue période d'observation pendant laquelle la comparabilité initiale s'estompe fréquemment.

	E	$\bar{E}$
M	a	b
$\bar{M}$	c	d

Les termes a, b, c, d représentent des effectifs. L'effectif total est (a + b + c + d).
Il faut remarquer que les taux d'exposé et de non exposés ne sont pas représentatif de la population par conséquent les rapports [a / (a + b)] et [c / (c + d)] n'ont pas de sens (double trait vertical rappelant cette notion).

Etude sur échantillon représentatif ; Enquête sur échantillon représentatif

Une telle étude est classée dans les enquêtes. Une enquête n'est jamais démonstrative. Après avoir défini la population d'étude, le tirage au sort sera effectué. On construira une table selon le schéma ci-dessous selon les facteurs maladie et exposition (facteur E supposé être lié à la maladie M). Ce type d'enquête permet de construire une table à quatre cellules contenant les effectifs pour les situations M E, $\bar{M}$ E, M $\bar{E}$, $\bar{M}$ E. Il est alors possible de déterminer s'il existe une liaison statistique entre les deux facteurs (calcul du Chi²) et calculer le risque relatif. La liaison obtenue n'est cependant pas une liaison de cause à effet.

	E	$\bar{E}$
M	a	b
$\bar{M}$	c	d

Les termes a, b, c, d représentent des effectifs. L'effectif total est (a + b + c + d).

Ce type d'enquête est souvent difficile à mettre en place et demande de grands effectifs lorsque la maladie est rare dans la population.

Etude pronostique

Formuler un pronostic, c'est estimer les probabilités des divers modes d'évolution d'une maladie. L'étude pronostique doit s'attacher à analyser l'évolution d'un groupe de sujet à partir d'une date ou d'un événement. Les études pronostiques s'appuient sur les études de cohorte et de survie.

Etude transversale ; Enquête transversale

C'est une étude dans laquelle l'exposition au facteur de risque et la maladie sont mesurées en même temps. Ce type d'étude permet de donner une vision à un instant donné. Elle permet d'obtenir la prévalence de diverses maladies.
Elle a un certains nombre d'inconvénients, en particulier elle ne permet pas d'apprécier la séquence "Exposition au risque ------> maladie"

Intervalle de confiance

Cette notion s'adresse aux grandeurs estimées à partir d'un échantillon. Nous considérons ici uniquement la fréquence d'un caractère ou la valeur moyenne d'une grandeur. La statistique inductive nous donne des informations sur la population à partir des informations obtenues sur l'échantillon. L'estimation d'une grandeur à partir d'un échantillon est entachée d'une incertitude puisque d'autres échantillons issues de la même population donneraient d'autres estimations. Lorsqu'un certain nombre de conditions sont réunies (population gaussienne ou effectif supérieur à 30) on peut trouver un intervalle tel qu'il y a 95% de chances pour que la vraie valeur du paramètre de la population (fréquence du caractère dans la population ou moyenne de la grandeur étudiée dans la population) soit dans l'intervalle. Cet intervalle est appelé "intervalle de confiance".

Médicament

Selon l'OMS (Organisation Mondiale de la Santé), est un médicament toute substance ou composition administrée pour guérir ou prévenir une maladie ou un symptôme, établir un diagnostic, modifier ou restaurer les fonctions de l'organisme.

Méthode expérimentale

L'expérience scientifique est une comparaison où tous les facteurs confondants (par ex. pour l'essai clinique : régression à la moyenne, effet placebo, évolution spontanée, terrain pathologique particulier, précision des instruments de mesure des critères de jugement), connus ou inconnus, jouent de la même manière sur les termes de la comparaison.
L'hypothèse testée sur laquelle s'appliqueront les tests statistiques, doit avoir été parfaitement formulée avant le début de l'expérience.

L'essai contrôlé est l'application à la recherche clinique des principes de la méthode décrite par Claude Bernard.

Mesures répétées

Cette notion mérite une clarification car il faut distinguer un protocole aboutissant à une analyse de la variance à mesures répétées permettant de déterminer une interaction entre deux facteurs et une analyse de la variance pour mesures répétées.
Dans le premier cas, mesures répétées veut dire que les mesures pour deux modalités des facteurs étudiées sont effectuées sur plusieurs individus alors que dans la deuxième situation,

les mesures sont répétées sur le même individu (par exemple à des temps différents ou à des doses différentes). La deuxième situation demande une analyse particulière car les mesures ne sont pas indépendantes (Analysis of variance for repeated measures).

Méthodes multivariées

Les modèles étudiés en recherche clinique sont complexes, une grandeur dépend en général de plusieurs facteurs. On cherche à effectuer une analyse "globale" en expliquant une variable (variable à expliquer) en fonction de plusieurs autres variables (variables explicatives). De telles méthodes sont appelées multivariées, nous citons par exemple la régression multiple, la régression logistique, le modèle de Cox pour l'étude des facteurs pronostics dans la survie, etc...

Ces méthodes sont très accessibles aujourd'hui et doivent être utilisées lorsque cela s'avère utile. Elles doivent cependant faire suite aux méthodes univariées qui permettent de comprendre plus facilement les relations pouvant exister entre grandeurs.

Il faut remarquer que dans toute étude on cherche à approcher la réalité médicale et biologique à travers des modèles définis. Il faut s'assurer que le modèle utilisé est capable de décrire le phénomène et par conséquent en valider son utilisation. Ceci est une démarche qui reste difficile car de nombreux facteurs tels que le modèle lui-même (additif, multiplicatif) et le codage des variables peuvent intervenir.

Moyenne

Il s'agit de la moyenne arithmétique définie comme la somme des valeurs divisée par l'effectif. Cette grandeur est le paramètre majeur caractérisant un ensemble de mesure ou plus classiquement une distribution. Cette valeur a son importance puisque dans les modèles concernés par la recherche clinique on recherche une différence de valeur moyenne entre deux ou plus de deux groupes lorsque la variable est quantitative. Le test statistique est construit pour tester une différence entre les valeurs moyennes des groupes.

Nombre de sujets nécessaire

L'objectif d'un essai thérapeutique est de conclure à une différence entre les effets mesurés dans les deux groupes au seuil de risque $\alpha = 0,05$ de conclure à tort que la différence existe. Ne pas conclure lorsque le test est non significatif peut signifier que la différence n'existe pas ou que le test manque de puissance. Pour éviter cette situation, il faut contrôler la puissance statistique du test par un calcul des effectifs nécessaires.

Le calcul du nombre de sujets nécessaire est basé sur les paramètres suivants :

- le risque α (de conclure à une différence qui n'existe pas) accepté (le plus souvent à 0,05, jamais au-dessus) ;
- le risque β (de ne pas saisir une différence alors qu'elle existe) accepté (le plus souvent à 0,10 ou 0,20 (ce qui correspond à une puissance statistique de 90 ou 80%) ;
- la différence jugée intéressante sur le plan clinique entre l'effet thérapeutique observé dans chacun des deux groupes ;
- la variabilité du critère de jugement (d'efficacité) principal, autrement dit la dispersion des valeurs mesurées sur l'échantillon.

Remarques :

- L'importance de définir un objectif principal, impliquant l'utilisation d'un critère de jugement principal pour sa mesure, apparaît parfaitement : le calcul du nombre de sujets nécessaire sera fait selon la variabilité de cette mesure (et non de celle d'autres critères de

jugement ou d'efficacité, dits alors secondaires) et la puissance de l'étude en découle directement.

- Il peut arriver que la dispersion des mesures du critère de jugement principal ne soit pas connue lors de la mise au point du protocole. Deux solutions sont alors possibles :
  - un calcul a posteriori vérifiant à partir des mesures effectuées que le nombre de sujets inclus dans l'étude permet d'obtenir une puissance statistique suffisante ;
  - la réalisation d'une pré-étude permettant d'estimer grossièrement cette variabilité.

- Il faut se méfier d'un excès d'optimisme dans la différence à mettre en évidence car une différence surestimée risque de ne pas permettre à terme de conclure par manque de puissance.

- Le calcul des effectifs nécessaires demande la connaissance de certaines grandeurs assez faciles à obtenir dans les cas de comparaisons entre deux groupes, mais beaucoup plus difficile à obtenir dans des plans d'étude complexe, comme par exemple dans l'analyse de la variance à plusieurs facteurs ou de nombreuses grandeurs sont à connaître.

- Une fois le calcul du nombre de sujets fait, il est habituel d'augmenter le effectifs de 10 ou 15% de façon à compenser les éventuelles sorties d'étude.

Observance (en anglais : compliance)

Degré de coïncidence entre le comportement d'une personne et les conseils et prescriptions donnés par son médecin.

Odds ratio

Cette grandeur, encore appelée rapport des cotes, est une valeur approchant le risque relatif lorsque le phénomène étudié est rare dans la population. C'est une grandeur qui contourne l'impossibilité de calculer le risque relatif lorsque l'étude est du type "Cas - Témoins".

	E	$\bar{E}$
M	a	b
$\bar{M}$	c	d

Les termes a, b, c, d représentent des effectifs. L'effectif total est (a + b + c + d).
On peut montrer que l'Odds Ratio (le rapport des cotes) noté OR = (a * d) / (b * c).
Cette grandeur est calculée dans divers tests utilisés en épidémiologie, en particulier elle constitue la grandeur principale calculée dans la régression logistique. Il faut néanmoins se rappeler que l'Odds Ratio ne tend vers le risque relatif que lorsque la maladie est rare dans la population.

Perdus de vue

Malades "disparus" avant la fin de l'essai et pour lesquels aucune information n'est plus disponible depuis cette "disparition". Cette situation est différente de celle où le malade reste accessible mais pour lequel manquent certaines informations (valeurs manquantes). Un pourcentage de ≥ 5% de perdus de vue dans un essai est inacceptable.

Petits effectifs

La notion de petits effectifs est associée à un calcul de chi² dans lequel on considère qu'il faut des effectifs théoriques au moins égaux à 5 pour être dans des conditions acceptables de calcul. Lorsque les effectifs théoriques sont plus petits que 5 on considère que l'on est dans le cas des petits effectifs. On utilise alors le calcul des probabilités exactes de Fisher pour pallier cette difficulté. Ce calcul est "exact" c'est à dire qu'il ne fait pas appel à des méthodes asymptotiques. Des méthodes ont été développés pour étendre le calcul des probabilités exactes de Fisher à des situations diverses, telles la recherche de liaisons dans des tableaux de contingences de dimension M*N (voir logiciel Statxact).

Il faut remarquer que ces calculs sont tout à fait utilisables en recherche clinique mais que le résultat est peu stable au regard de la taille de l'échantillon, les conclusions sont donc très fragiles.

Plan d'expérience (plan expérimental)

Les plus courants sont :

- **plan en groupes parallèles :**

Après une période de pré-inclusion, les participants sont randomisés en 2 (ou plus) groupes dont chacun va recevoir un des traitements testés.

- **plan en chassé-croisé (ou cross-over) :**

Chaque sujet est son propre témoin : il reçoit un traitement A pendant une première période et l'autre traitement B pendant une deuxième période, l'allocation aléatoire s'appliquant à l'ordre A-B ou B-A de sorte qu'autant de patients recevront d'abord B ou A.

- **plan factoriel :**

Utilisé pour évaluer la contribution de chacun des traitements prescrits en association. Par exemple, un plan factoriel conçu pour évaluer l'effet antihypertenseur d'une association de bêta-bloquants et de diurétiques comportera 4 groupes de traitements (parallèles) en double insu : association bêta-bloquant + diurétique ; bêta-bloquant + placebo du diurétique ; diurétique + placebo du bêta-bloquant ; placebo du bêta-bloquant + placebo du diurétique.

Principe du test statistique

La comparaison de deux traitements est réalisée par une stratégie définie dans le principe du test statistique. En comparant deux échantillons, par exemple un échantillon traité par A et un échantillon traité par B, il s'agira de dire s'ils proviennent de la même population ou non (ou que les deux traitements ne montrent pas de différence ou non). Cette comparaison est effectuée à l'aide de mesures issues de chaque échantillon, une fréquence ou une moyenne par exemple.

Dans le principe du test statistique, on pose une hypothèse H_0 formulant que les deux échantillons proviennent de la même population. On calcule ensuite une grandeur discriminante à partir des différences constatées sur les grandeurs étudiées.

Lorsque la différence constatée est grand, on rejette l'hypothèse H_0 et l'on admet que les deux échantillons ne proviennent pas de la même population, par conséquent que A est différent de B. La statistique nous apprend que cette conclusion est effectuée sous le risque $\alpha = 0,05$ d'avoir rejeté H_0 à tort.

Lorsque cette différence est trop faible on garde l'hypothèse H_0 et l'on conclu que les deux échantillons proviennent de la même population, on considère qu'aucune différence significative entre A et B n'a été mise en évidence. La statistique nous apprend que cette

conclusion est effectuée sous le risque β de ne pas avoir mis la différence en évidence. Ce risque peut être diminué en augmentant les effectifs.

Puissance d'un essai clinique

C'est l'aptitude d'un essai à mettre en évidence l'effet du médicament testé. Cette puissance relève de quatre composantes :

- **Puissance statistique :**

Le nombre de malades inclus est un élément essentiel de la puissance d'un essai clinique. La puissance statistique (1-β) est le complément du risque de ne pas saisir une différence qui existe (risque β ou de 2^{e} espèce). Le calcul du nombre de sujets et sa réalisation dans l'étude sont indispensables pour assurer cette puissance statistique.

- **Puissance pharmacologique :**

Elle est inhérente à la logistique de l'essai. Ainsi, il faut que le médicament à tester et médicament de référence soient administrés à posologies adéquates, que l' observance des traitements soit satisfaisante, que la fréquence des prises soit adaptée à la pharmacocinétique des médicaments testés.

- **Puissance méthodologique :**

Le plan d'expérience doit être adapté à la situation explorée : par ex. un essai en chassé-croisé ne peut être utilisé lorsque la pathologie traitée varie spontanément et rapidement. La qualité de la réalisation de l'étude intervient également : sorties d'études, valeurs manquantes, participants perdus de vue, sont autant de paramètres susceptibles de réduire la puissance d'une étude.

- **Puissance analytique :**

La méthode statistique doit être adaptée au protocole choisi et aux données recueillies. Ainsi, l'utilisation répétée des mêmes variables pour des comparaisons multiples demande une adaptation du seuil du risque α et entraîne une diminution de la puissance éventuellement compensée par une augmentation des effectifs de l'échantillon.

Randomisation

C'est une méthode pour diviser une population en groupes de telle manière que l'affectation des sujets laisse les groupes comparables. Chaque sujet a les mêmes chances d'être affecté à l'un ou l'autre des groupes. La randomisation permet d'éviter le biais d'attribution et par conséquent de conclure à l'action du traitement lorsqu'une différence significative apparaît. Lors d'une randomisation, les allocations d'individus doivent être définies à l'avance. On pourra se servir de tables de randomisation pour affecter les individus aux divers groupes (dans le cas de deux groupes l'un des groupes recevra le traitement A, l'autre le traitement B).

Dans certains cas on peut utiliser des méthodes permettant d'équilibrer les effectifs tous les n patients (randomisation par blocs) ou randomiser pour représenter de façon équilibrée un paramètre dans les groupes (par exemple l'âge) en randomisant par stratification.
Comparer des patients recevant un nouveau traitement à ceux recevant un ancien n'est pas une étude randomisée, il est presque certain que des biais sont présents dans l'étude.

Référence (médicament ou traitement de)

C'est le traitement du groupe servant de contrôle. Son choix dépend de la situation thérapeutique étudiée :

- il n'existe aucune thérapeutique pour la maladie considérée : placebo ;
- il existe un traitement établi (soit par des essais contrôlés antérieurs, soit par une longue et indiscutable expérience clinique) : médicament validé à la posologie active, ou, si l'éthique le permet, un placebo.

Régression à la moyenne

Phénomène statistique obligatoire, dont l'intensité est liée à la variabilité intra-sujet de la variable considérée, observé chaque fois qu'est suivie l'évolution d'un paramètre au sein d'un échantillon tronqué (i.e. limité par un seuil) d'une population.

Ainsi, l'identification de personnes hypertendues dans une population consiste à employer un seuil de pression artérielle. Lors des répétitions des mesures, la moyenne des valeurs de pression artérielle des sujets ainsi sélectionnés lors d'une première mesure va se rapprocher de la moyenne des valeurs de la PA de l'ensemble de la population. La régression à la moyenne est d'autant plus forte que le seuil choisi est plus extrême. Ceci explique des erreurs d'analyse, telles que par exemple : l'effet du traitement est d'autant plus marqué que les taux de départ sont plus élevés.

Risque alpha ou risque de 1re espèce

C'est le risque, accepté dans une étude, de conclure à une différence significative alors qu'en réalité elle n'existe pas. Ce risque est fixé arbitrairement fixé à 5% en médecine et en biologie.

Risque Bêta ou risque de 2e espèce

C'est le risque, accepté dans une étude, de ne pas conclure à une différence significative alors qu'elle existe. Ce risque peut être diminué en augmentant les effectifs de l'étude.

Risque relatif

Considérons deux groupes de sujets dont l'un est exposé à une facteur censé favoriser une maladie, l'autre n'est pas exposé. On souhaite savoir si l'exposition au facteur engendre davantage de maladies que la non exposition. Pour comparer les deux situations, on calcule "le risque relatif". Le risque relatif est le rapport entre le taux d'incidence chez les exposés par le taux d'incidence chez les non exposés.

	E	$\bar{E}$
M	a	b
$\bar{M}$	c	d

Les termes a, b, c, d représentent des effectifs. L'effectif total est (a + b + c + d).
Le risque relatif, noté RR = [a / (a + c)] / [b / (b + d)]
Ce risque peut être calculé par une étude de cohorte ou une étude sur échantillon représentatif.
Les termes a / (a + c) et b / (b + d) représentent les risques absolus (taux d'incidence) d'être atteint par la maladie dans l'un et l'autre groupe.

Le risque attribuable est noté RA, c'est le taux d'incidence chez les exposés - le taux d'incidence chez les non exposés.
Tous ces termes sont des estimations obtenues à partir d'un échantillon d'individus, il convient dès lors de leur associer un intervalle de confiance.

Sensibilité

La sensibilité est une notion qui caractérise les qualités diagnostiques d'un signe clinique (S), d'un test clinique ou paraclinique (T). La sensibilité est la probabilité conditionnelle de la présence du signe S^+ ou du test positif T^+ lorsque la maladie est présente (M^+).

	M^+	M^-
S^+ ou T^+	a	b
S^- ou T^-	c	d

La sensibilité = [a / (a + c)] = Probabilité (S^+ / M^+) ou encore Probabilité (T^+ / M^+)
L'effectif a correspond aux vrais positifs, l'effectif c correspond aux faux négatifs

Simple – insu

Etude contrôlée dans laquelle le patient ne connaît pas dans quel groupe de traitement il est affecté (il ignore quel traitement lui est attribué). Cette méthode a pour objectif d'éviter les biais liées aux réponses du patient.

Spécificité

La spécificité est une notion qui caractérise les qualités diagnostiques d'un signe clinique (S), d'un test clinique ou paraclinique (T). La spécificité est la probabilité conditionnelle de l'absence du signe S ou de la négation du test T lorsque la maladie est absente (M^-).

	M^+	M^-
S^+ ou T^+	a	b
S^- ou T^-	c	d

La spécificité = [d / (b + d)] = Probabilité (S^- / M^-) ou encore Probabilité (T^- / M^-)
L'effectif d correspond aux vrais négatifs, l'effectif b correspond aux faux positifs.

Statistiques

Méthodes essentiellement descriptives centrées sur le dénombrement.

Statistique

Méthodes construites autour de l'organisation d'une étude, des lois de probabilités et des tests pour apporter une validation scientifique à une étude.

Survie

L'étude de la survie fait appel à plusieurs modèles intégrant plus ou moins l'ensemble de l'information connue. La méthode de Kaplan Meier utilise le maximum d'informations puisqu'elle tient compte de chaque événement. Le modèle de Cox permet l'étude des facteurs pronostiques.

Tests non paramétrique

Méthodes statistiques permettant la comparaison de deux ou plusieurs groupes sans faire d'hypothèses sur la distribution des grandeurs étudiées. Ces méthodes sont basées sur des calculs de probabilités. Les tests non paramétriques doivent être utilisés lorsque les conditions nécessaires à l'utilisation des tests paramétriques ne sont pas satisfaites. Ces tests sont en général moins puissants que les tests paramétriques.

Tests paramétriques

Méthodes statistiques utilisées lorsque les lois de distribution sont connues et que celle-ci sont caractérisées par leurs paramètres principaux (moyenne, écart-type).

Valeurs prédictives positives et négatives

Ces grandeurs mesurent la performance d'une stratégique diagnostique (d'un test).

La Valeur Prédictive Positive (VPP) est la proportion de tests positifs qui correspond à de vrais malades. C'est le rapport du nombre de vrais positifs sur le nombre total de test positifs. La VPP correspond à la probabilité pour le patient d'être malade lorsque le test est positif.
VPP = a / (a+b) = Probabilité (M^+ / S^+)

	M^+	M^-
S^+ ou T^+	a	b
S^- ou T^-	c	d

La Valeur Prédictive Négative (VPN) est la proportion des tests négatifs qui correspond à des non-malades. C'est le rapport du nombre de vrais négatifs sur le nombre total de tests négatifs. La VPN correspond à la probabilité pour le patient de ne pas être malade lorsque le test est négatif.
VPN = d / (c+d) = Probabilité (M^- / S^-)