

LECTURE CRITIQUE D'UN ARTICLE MEDICAL

Support de cours pour la préparation
de l'Examen National Classant

UFR Médecine Montpellier-Nîmes
2011-2012

Liste des auteurs

Dorine Neveu, MCF

Pierre Dujols, PU-PH

Pascale Fabbro-Peray, MCU-PH

Nous remercions les membres de l'équipe pédagogique de lecture critique d'articles pour leur contribution et la relecture du document.

Fabienne Séguret, PH

Jean-Luc Faillie, AHU

Sophie Bastide, AHU

Grégoire Mercier, AHU

Nicolas Molinari, MCU-PH

Thibault Mura, AHU

Nicolas Nagot, MCU-PH

Marie-Christine Picot, PH

Sommaire

INTRODUCTION	5
ABREVIATIONS	6
OBJECTIF 1 : SAVOIR IDENTIFIER L'OBJET D'UN ARTICLE MEDICAL SCIENTIFIQUE	7
OBJECTIF 2 : SAVOIR IDENTIFIER LA QUESTION ETUDIEE	8
OBJECTIF 3 : IDENTIFIER LES CARACTERISTIQUES (DONNEES DEMOGRAPHIQUES) DE LA POPULATION ETUDIEE, A LAQUELLE LES CONCLUSIONS POURRONT ETRE APPLIQUEES ET OBJECTIF 4 : ANALYSER LES MODALITES DE SELECTION DES SUJETS, CRITERES D'INCLUSION ET CRITERES DE NON-INCLUSION	10
OBJECTIF 5 : IDENTIFIER LA TECHNIQUE DE RANDOMISATION ET VERIFIER SA COHERENCE, LE CAS ECHEANT.	13
OBJECTIF 6 : DISCUTER LA COMPARABILITE DES GROUPES SOUMIS A LA COMPARAISON	16
OBJECTIF 7 : DISCUTER L'EVOLUTION DES EFFECTIFS ETUDIES ET LEUR COHERENCE DANS LA TOTALITE DE L'ARTICLE ; SAVOIR SI LE CALCUL DU NOMBRE DE SUJETS NECESSAIRES A ETE EFFECTUE A PRIORI	18
OBJECTIF 8 : S'ASSURER QUE LA METHODE EMPLOYEE EST COHERENTE AVEC LE PROJET DU TRAVAIL, QUE LA METHODOLOGIE EST EFFECTIVEMENT SUSCEPTIBLE D'APPORTER UNE REPONSE A LA QUESTION POSEE DANS L'INTRODUCTION.	19
OBJECTIF 9 : VERIFIER QUE LES ANALYSES STATISTIQUES (EN FONCTION DE NOTIONS ELEMENTAIRES) SONT COHERENTES AVEC LE PROJET DU TRAVAIL ; CONNAITRE LES LIMITES DE L'ANALYSE PAR SOUS GROUPE ; CONNAITRE LA NOTION DE PERDUS DE VUE.	23
OBJECTIF 10 : VERIFIER LE RESPECT DES REGLES D'ETHIQUE	27
OBJECTIF 11°: ANALYSER LA PRESENTATION, LA PRECISION ET LA LISIBILITE DES TABLEAUX ET DES FIGURES, LEUR COHERENCE AVEC LE TEXTE ET LEUR UTILITE.	29
OBJECTIF 12°: VERIFIER LA PRESENCE DES INDICES DE DISPERSION PERMETTANT D'EVALUER LA VARIABILITE DES MESURES ET LEURS ESTIMATEURS	31
OBJECTIF 13°: DISCUTER LA NATURE ET LA PRECISION DES CRITERES DE JUGEMENT DES RESULTATS	32
OBJECTIF 14°: RELEVER LES BIAIS QUI ONT ETE DISCUTES, RECHERCHER D'AUTRES BIAIS D'INFORMATION ET DE SELECTION EVENTUELS NON PRIS EN COMPTE DANS LA DISCUSSION ET RELEVER LEURS CONSEQUENCES DANS L'ANALYSE DES RESULTATS	35
OBJECTIF 15 : VERIFIER LA LOGIQUE DE LA DISCUSSION ET SA STRUCTURE, RECONNAITRE CE QUI RELEVE DES DONNEES DE LA LITTERATURE ET CE QUI EST L'OPINION PERSONNELLE DE L'AUTEUR	42

OBJECTIF 16°: DISCUTER LA SIGNIFICATION STATISTIQUE DES RESULTATS.	43
OBJECTIF 17° DISCUTER LA PERTINENCE CLINIQUE DES RESULTATS	45
OBJECTIF 18°: VERIFIER QUE LES RESULTATS OFFRENT UNE REPONSE A LA QUESTION ENONCEE ET OBJECTIF 19°: VERIFIER QUE LES CONCLUSIONS SONT JUSTIFIEES PAR LES RESULTATS.	48
OBJECTIF 20 : INDIQUER LE NIVEAU DE PREUVE DE L'ETUDE (GRILLE DE L'HAS / ANAES)	49
OBJECTIF 21°: DISCUTER LA OU LES APPLICATIONS POTENTIELLES PROPOSEES PAR L'ETUDE	50
OBJECTIF 22 : IDENTIFIER LA STRUCTURE IMRAD (INTRODUCTION, MATERIEL ET METHODE, RESULTATS, DISCUSSION) ET S'ASSURER QUE LES DIVERS CHAPITRES REPONDENT A LEURS OBJECTIFS RESPECTIFS.	52
OBJECTIF 23 : FAIRE UNE ANALYSE CRITIQUE DE LA PRESENTATION DES REFERENCES	54
OBJECTIF 24 : FAIRE UNE ANALYSE CRITIQUE DU TITRE	55
ADDENDUM 1°: BASES EPISTEMOLOGIQUES DE LA CONNAISSANCE BIOMEDICALE	56
ADDENDUM 2°: LES TYPES D'ETUDES	57
ADDENDUM 3°: EVALUATION D'UN TRAITEMENT	64
ANNEXE 1 : OUTILS STATISTIQUES EN FONCTION DE LA PROBLEMATIQUE	65
REFERENCES BIBLIOGRAPHIQUES	67

INTRODUCTION

Ce document est un support de cours de lecture critique d'article (LCA). Il vous donne une **méthode de travail** que vous mettrez en pratique dans les enseignements dirigés de LCA. Il fait appel à un raisonnement et doit être complété par vos connaissances cliniques et par les exemples vus en enseignements dirigés de LCA. Il fait appel à des notions définies dans le glossaire du CNCI que vous trouvez sur le site du CNCI : <http://www.cnci.univ-paris5.fr/medecine/>

Il traite les objectifs pédagogiques de l'épreuve de LCA formulés par le CNCI , listés ci-dessous.

Il ne traite pas spécifiquement de la méthodologie du résumé pour laquelle vous pouvez consulter le site du LIPCOM ou l'ENT.

SAVOIR IDENTIFIER

- 1) Savoir identifier l'objet d'un article médical scientifique parmi les suivants: évaluation d'une procédure diagnostique, d'un programme de dépistage, d'un traitement, estimation d'un pronostic, enquête épidémiologique
- 2) Savoir identifier la question posée par les auteurs (hypothèse)

SAVOIR CRITIQUER LA METHODOLOGIE

- 3) Identifier les caractéristiques (données démographiques) de la population étudiée, à laquelle les conclusions pourront être appliquées.
- 4) Analyser les modalités de sélection des sujets, critères d'inclusion et de non-inclusion et d'exclusion.
- 5) Identifier la technique de randomisation et vérifier sa cohérence, le cas échéant.
- 6) Discuter la comparabilité des groupes soumis à comparaison
- 7) Discuter l'évolution des effectifs étudiés et leur cohérence dans la totalité de l'article; savoir si le calcul du nombre de sujets nécessaires a été effectué à priori.
- 8) S'assurer que la méthode employée est cohérente avec le projet du travail, et qu'elle est effectivement susceptible d'apporter "une" réponse à la question posée dans l'introduction.
- 9) Vérifier que les analyses statistiques (en fonction des notions élémentaires) sont cohérentes avec le projet; connaître les limites de l'analyse par sous-groupes; connaître la notion de perdus de vue.
- 10) Vérifier le respect des règles d'éthique.

ANALYSER LA PRESENTATION DES RESULTATS

- 11) Analyser la présentation, la précision et la lisibilité des tableaux et des figures, leur cohérence avec le texte et leur utilité.
- 12) Vérifier la présence des indices de dispersion permettant d'évaluer la variabilité des mesures et de leurs estimateurs.

CRITIQUER L'ANALYSE DES RESULTATS ET DE LA DISCUSSION

- 13) Discuter la nature et la précision des critères de jugement des résultats.
- 14) Relever les biais qui ont été discutés. Rechercher d'autres biais d'information et de sélection éventuels non pris en compte dans la discussion et relever leurs conséquences dans l'analyse des résultats.
- 15) Vérifier la logique de la discussion et sa structure. Reconnaître ce qui relève des données de la littérature et ce qui est opinion personnelle de l'auteur.
- 16) Discuter la signification statistique des résultats.
- 17) Discuter la pertinence clinique des résultats.
- 18) Vérifier que les résultats offrent une réponse à la question annoncée.
- 19) Vérifier que les conclusions sont justifiées par les résultats.
- 20) Indiquer le niveau de preuve de l'étude (grille de l'HAS).

EVALUER LES APPLICATIONS CLINIQUES

- 21) Discuter la ou les applications potentielles proposées par l'étude

ANALYSER LA FORME DE L'ARTICLE

- 22) Identifier la structure IMRAD (Introduction, Matériel et méthode, Résultats, Discussion) et s'assurer que les divers chapitres répondent à leurs objectifs respectifs.
- 23) Faire une analyse critique de la présentation des références
- 24) Faire une analyse critique du titre.

Pour chaque objectif, on donne d'abord les questions essentielles à se poser, on localise ensuite la section dans l'article où il faut rechercher les éléments de réponse, enfin on fait des rappels théoriques relatifs à chaque objectif en donnant des exemples.

ABREVIATIONS

AUC : aire sous la courbe

CJP : critère de jugement principal

CJS : critère de jugement secondaire

HR : hazard ratio

IC95% : intervalle de confiance à 95%

NSN : nombre de sujets nécessaire

OR : odds ratio

RAR : réduction absolue de risque

ROC : receiver operating curve

RR : risque relatif

RRR : réduction relative de risque

OBJECTIF 1 : SAVOIR IDENTIFIER L'OBJET D'UN ARTICLE MEDICAL SCIENTIFIQUE

1. LES QUESTIONS A SE POSER

L'objet de la publication est-il:

- L'évaluation d'une procédure diagnostique
- L'évaluation d'un programme de dépistage
- L'évaluation d'un traitement
- L'estimation d'un pronostic
- Une enquête épidémiologique

Ces 5 objets étant les cadres des études sur lesquelles peut porter l'épreuve de l'ECN

2. OU CHERCHER

Dans le titre, en fin de section introduction, dans la section méthode.

L'objet sera déduit de la question étudiée et du schéma de l'étude (encore appelé schéma expérimental). Ces trois points, objet - question - schéma, devront être cohérents.

3. RAPPELS THEORIQUES :

L'**objet d'un article** est, pour l'épreuve de l'ECN, le **grand cadre de l'étude** qui a donné lieu à la publication analysée.

4. ATTENTION :

Faire attention: l'objet n'est pas le plan expérimental, même si la connaissance du plan expérimental est un des éléments qui permettra, avec la question étudiée de déterminer l'objet.

Exemple : Estimation d'un pronostic : cet objet s'applique à toute étude portant sur des sujets malades dont on veut déterminer le pronostic en termes de survie, rechute, rémission, guérison

OBJECTIF 2 : SAVOIR IDENTIFIER LA QUESTION ETUDIEE

1. LES QUESTIONS A SE POSER

- a. La question, énoncée dans l'objectif (les objectifs), est-elle une question de recherche?
- b. La question est-elle énoncée dans toutes ses composantes ?

2. OU CHERCHER

Pour a: dans l'introduction

Pour b: en fin de section Introduction et, parfois, début de section Méthodes et/ou paragraphe considérations statistiques

3. RAPPELS THEORIQUES :

Il faut tout d'abord comprendre et se souvenir qu'une question étudiée dans une publication:

- A. Doit être une vraie question de recherche (cf ci-dessous §A) donnant une avancée des connaissances / confirmation scientifique
- B. Doit être bien énoncée, c'est à dire avec toutes ses composantes (cf ci-dessous §B)
- C. Doit être hiérarchisée
- D. Est la pierre angulaire de l'étude. La méthode doit permettre de répondre à la question. Les conclusions, et les éventuelles recommandations se rapportent à la question principale.
- E. Dont l'Observé (le phénomène que l'on étudie) est mesurable par un critère de jugement adéquat selon un Schéma d'Etude adapté.

A. Ce qu'est une question de recherche

Toute question bio-médicale n'est pas une question de recherche.

Une question de recherche doit être pertinente et argumentée, c'est à dire:

- Bien énoncée (cf §B)
- Face à une problématique d'intérêt, argumentée sur l'état actuel des connaissances, s'appuyant sur des références publiées, portant sur les 3 points suivants (**SCL**)
 - ✓ **Santé Publique** (fréquence de la maladie / du problème en santé, augmentation de fréquence, coût pour la société, mortalité importante ...)
 - ✓ **Clinique ou physiopathologique** (pronostic grave ou invalidant, séquelles fréquentes / importantes, difficulté du diagnostic par les moyens habituels ...)
 - ✓ **Lacunes dans les connaissances** (mécanismes physiopathologiques, arguments génétiques) ou limites dans les études précédentes de même objet.

B. Composantes d'une question de recherche

Les composantes d'une question de recherche sont (**PICO**):

- **Population** : ensemble d'unités, le plus souvent des personnes définies sur des critères précis
- **Intervention** : facteur étudié (action, programme, facteur ou thérapie que le groupe de comparaison reçoit ou auquel il est exposé)
- **Comparateur** : action, facteur, programme ou thérapie à qui on compare l'intervention et auquel le groupe de comparaison, quand il existe, est exposé.
 - Exemples
 - un traitement « classique » ou l'absence de traitement, ou un placebo dans un essai médicamenteux
 - la « non exposition » au facteur de risque dans une enquête épidémiologique analytique
- **Observé (Outcome en anglais)** : phénomène d'intérêt (clinique, physiopathologique, ...), sur lequel l'intervention est supposée avoir un impact.

Ce phénomène sera mesuré par 'un critère de jugement ou d'évaluation (ex: le diabète est mesuré au travers de la glycémie) (cf §C et objectif 13)

Exemple : Comparer l'efficacité (O) et la tolérance (O) de l'énoxaparine (I) et de l'héparine non fractionnée (C) chez des patients avec un syndrome coronarien aigu (P)

Selon la recherche, certains des éléments constitutifs de la question sont sans objet.

Exemple : dans les enquêtes descriptives, les composantes C et I sont absentes. Estimer la prévalence des infections nosocomiales (O) parmi les patients des établissements hospitaliers français publics et privés (P)

C. Mesure de l'observé

La mesure de l'Observé est effectuée:

- Par un **Critère de Jugement** qui est le paramètre que l'on va mesurer pour représenter l'Observé. Ce critère, aussi appelé Critère d'Evaluation, sert au calcul du nombre de sujets nécessaires à l'étude [cf Objectif-LCA n° 13]
- Selon un **Schéma d'Etude**, ou type d'étude (study design) [cf Objectif-LCA n° 8]
- *Remarque : parfois l'expression "plan expérimental" est utilisée pour le schéma de l'étude. Cette expression ne signifie pas que l'étude est expérimentale, que l'intervention/facteur étudié est contrôlée.*

4. ATTENTION :

1. Toutes les composantes de la question ne sont pas obligatoires selon les objets d'étude [cf Objectif-LCA n°1]
2. Si l'observé est directement mesurable, il peut être aussi le critère de jugement (exemple: survie).
3. Une étude ne devrait correspondre qu'à une question et donc un critère de jugement, un résultat et une conclusion. Cependant, il arrive souvent que plusieurs questions soient posées dans une étude. On parlera alors d'objectif principal et objectifs secondaires. **L'objectif principal** (= la question principale) est celle qui a été à la base de l'étude et **fonde le résultat principal de l'étude**. Elle est énoncée en premier. Elle a permis (quand licite) de calculer le nombre de sujets nécessaires. Les **objectifs secondaires**, questions secondaires auxquelles l'étude essaie de répondre, **ont une valeur surtout exploratoire** et doivent être peu nombreux. Comme ils n'ont pas participé au calcul du nombre de sujets nécessaires, les analyses faites sur leurs mesures risquent de ne pas avoir une puissance suffisante pour une conclusion étayée.
4. **Ne pas confondre objectif et hypothèse**. Dans l'hypothèse, on formule la réponse escomptée. Dans l'objectif, on reste neutre.

- **Exemple:**

Objet de l'article: Evaluation de l'efficacité thérapeutique d'une intervention médicale

Exemple d'objectif : Evaluer l'efficacité d'un médicament antivitaminé K pour prévenir la survenue d'accident thrombo-embolique chez des sujets de plus de 65 ans, en arythmie cardiaque par fibrillation atriale.

Hypothèse : les antivitaminés K diminuent le risque de survenue d'accident thrombo-embolique chez des sujets de plus de 65 ans, en arythmie cardiaque par fibrillation atriale.

OBJECTIF 3 : IDENTIFIER LES CARACTERISTIQUES (DONNEES DEMOGRAPHIQUES) DE LA POPULATION ETUDIEE, A LAQUELLE LES CONCLUSIONS POURRONT ETRE APPLIQUEES ET OBJECTIF 4 : ANALYSER LES MODALITES DE SELECTION DES SUJETS, CRITERES D'INCLUSION ET CRITERES DE NON-INCLUSION

1. LES QUESTIONS A SE POSER

- a. A quelle population souhaite-on a priori proposer l'intervention si elle s'avère efficace et bien tolérée ou extrapoler les résultats? (**population cible**)
- b. A partir de quelle population a-t-on extrait le (les) échantillon(s) de l'étude ? (**population source**)
- c. Le mode de recrutement des sujets est-il précisé ?
- d. Le lieu de l'étude est-il précisé ?
- e. Les critères d'inclusion et de non inclusion des sujets sont-ils précisés ?
- f. L'échantillon est-il représentatif de la population source ?
- g. La base de recrutement (sondage) permet-elle d'atteindre toute la population définie par les critères d'inclusion et de non inclusion ?
- h. A quelle population les conclusions peuvent-elles être appliquées ?

2. OU CHERCHER

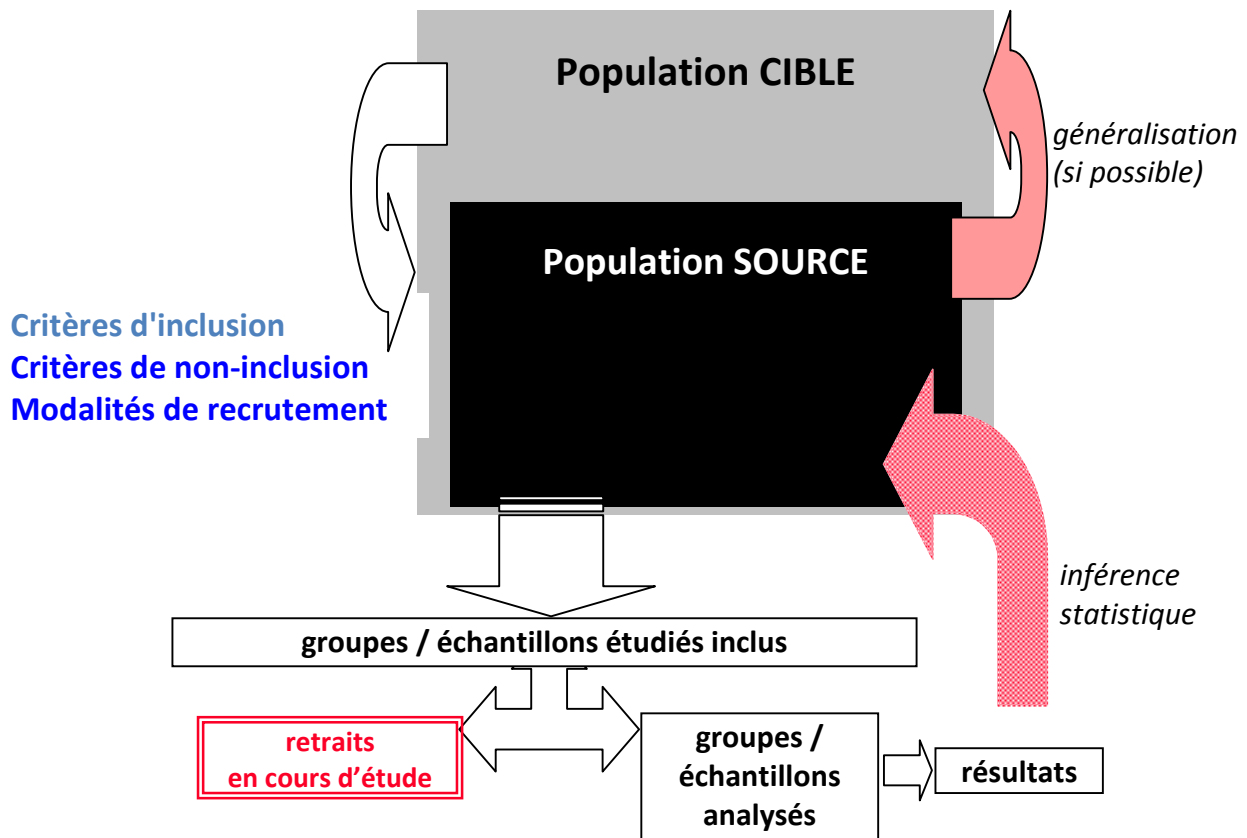
Population cible: dans le titre, à la fin de l'introduction (la question), au début de la partie méthodes, au début de la partie discussion

Population source: partie méthodes

Population à laquelle les conclusions peuvent être appliquées : méthode, résultats, diagramme de flux, fin de la partie discussion

3. RAPPELS THEORIQUES

Toute population incluse dans une étude a des caractéristiques spécifiques, qu'il est important de décrire, en particulier pour interpréter les résultats de l'étude et leur caractère généralisable (validité externe).



Population-Cible: population à laquelle les résultats d'une étude pourront a priori être étendus.

Population-Source: population au sein de laquelle l'échantillon a été tiré Avec le lieu d'étude elle permet la reproductibilité de l'étude (possible en prospectif mais plus difficile en rétrospectif).

2 niveaux : critères d'inclusion et de non inclusion, et modalités de recrutement (ex: *listes électorales ...*).

Modalités de recrutement des sujets (critères d'inclusion et de non inclusion) puis modalités de sélection des sujets analysés permettent de déterminer les individus éligibles pour l'étude et d'argumenter à quelle population on peut généraliser des résultats (population cible ou une autre à définir). Utiliser pour cela le diagramme de flux.

A. Modalités de recrutement

Qui : malades, sujets sains, volontaires sains ou malades, cotisants d'assurance sociale ...

Comment : par publicité, sur liste électorale, sur liste d'individus présents

Où : un/des service(s) spécialisé(s), un hôpital, des hôpitaux, médecine de ville, entreprise, école, ville, région ..., niveau national ou international ...

Quand : *date*

Il convient de regarder si les modalités de recrutement ne ciblent pas des sujets à profil particulier donc non représentatifs de la population cible. Si les modalités de recrutement ne permettent pas d'atteindre toute la population source, la généralisation est problématique.

B. Lieu(x) ou site(s) d'étude

Le(s) lieu(x) où l'étude s'est déroulée est un élément important de l'évaluation de la représentativité de l'échantillon et de l'applicabilité des résultats de l'étude dans le contexte particulier du lecteur.

Il est en particulier important de savoir si l'étude:

- s'est déroulée dans un centre de référence (plateau technique sophistiqué) ou dans les hôpitaux généraux ...
- portait sur des soins de santé primaires (en médecine ambulatoire)
- portait sur des non malades (essais d'intervention), ciblés ou non ...
- se déroulait dans un lieu géographique particulier (avec population particulière) ...
- était mono ou multicentrique

C. Critères d'inclusion des sujets

Définition: Ensemble de critères qui définissent de façon précise les caractéristiques des sujets qui peuvent entrer dans une étude.

Ils reflètent la population cible de façon positive. Ils peuvent être:

- des critères sociodémographiques: origine ethnique, origine sociale, âge, sexe ...
- des critères géographiques: origine, habitation, hospitalisation ...
- des critères cliniques

D. Critères de non-inclusion des sujets

Définition: ensemble des critères faisant que les sujets ne peuvent pas être inclus dans une étude ou un essai.

Ce sont des **critères de limitation de la sélection**.

Ils appartiennent à une des 3 classes suivantes:

- raison de prudence (prévention d'événements indésirables)
 - ✓ Contre-indication à l'intervention étudiée (grossesse, autre traitement, maladies associées...)
 - ✓ Contre-indication à une exploration nécessaire au critère de jugement
- difficulté potentielle d'évaluation des critères de jugement (on veut que l'effet mesuré dans chaque groupe via le critère de jugement soit dû à la pathologie et/ou au traitement et non à une co-morbidité ou à un traitement concomitant)
 - ✓ Traitement associé interférent (interactions médicamenteuses, ...)

- ✓ Maladie associée ou handicap pouvant fausser les évaluations
- difficultés potentielles de suivi
 - ✓ Pathologie associée prioritaire dans les soins
 - ✓ Motivation insuffisante
 - ✓ Possibilité de perte de vue
 - ✓ Co-morbidités ou caractéristiques favorisant une mauvaise observance ou un abandon d'étude

E. Population à laquelle les conclusions peuvent être appliquées

- **Population cible** (généralement caractérisée par les critères d'inclusion).
 - Si l'échantillon est représentatif de la population définie par les critères d'inclusion et de non inclusion (pas de biais majeur de sélection lié à des exclusions des analyses, des perdus de vue, des données manquantes, des modalités de recrutement).
 - Et si le diagramme de flux est donné avec le processus de sélection entre la pré-sélection (pré-sélection des sujets selon les critères d'inclusion) et l'inclusion,
 - Si le taux de sujets non inclus parmi les pré-sélectionnés est peu élevé et les motifs de non-inclusion (refus de participer, critères de non inclusion restrictifs) suggèrent que ces non-inclus sont comparables aux sujets analysés. (Chercher aussi dans les sections résultats et discussion des informations complémentaires.)
 - Attention : Si ce taux de sujets non inclus est élevé (de l'ordre de 20-30%), la généralisation des conclusions est problématique. Argumenter en vous basant sur les motifs de non inclusion.
- **Population source** (définie par les modalités de recrutement, les critères d'inclusion et de non inclusion)
 - Si échantillon analysé représentatif de la population source.
 - Attention aux modalités de recrutement.

ATTENTION SI ETUDE AVEC ANALYSES EN SOUS-GROUPES. CF § OBJECTIF 9.

OBJECTIF 5 : IDENTIFIER LA TECHNIQUE DE RANDOMISATION ET VERIFIER SA COHERENCE, LE CAS ECHEANT.

1. LES QUESTIONS A SE POSER

- L'étude devrait-elle être randomisée ?
- Si oui, l'étude comporte elle une randomisation ?
- Si oui, la méthode est-elle décrite ?
- Si oui, la randomisation est-elle de qualité c'est à dire garantit-elle l'imprévisibilité du traitement qui sera alloué au moment de l'inclusion, (allocation du traitement non prédictible, avec des modalités adaptées à l'objectif de l'étude, à un moment adéquat dans le déroulement de l'étude) ? = respect de la clause d'ignorance
- Si oui, le mode de randomisation est-il particulier ?
- Si l'essai clinique est multicentrique, la randomisation est-elle stratifiée sur les centres ?
- Si des analyses en sous-groupes sont prévues, la randomisation a t-elle été stratifiée sur les variables qui définissent les sous-groupes ?

2. OU CHERCHER

Dans la partie méthode pour la description du mode et du moment de la randomisation

(Les auteurs de l'article doivent donner suffisamment d'information afin que le lecteur puisse s'assurer que la méthode employée est bien une allocation au hasard et qu'il n'y a pas eu de biais dans l'allocation des sujets)

3. RAPPELS THEORIQUES :

A. Définitions

Randomisation : tirage au sort des patients permettant une répartition au hasard, aléatoire, des patients dans deux ou plusieurs groupes (*définition du CNCI*).

Attention :

Le terme randomisation est consacré à un tirage au sort qui aboutit à la constitution de plusieurs groupes.

Quand on a un seul groupe, on utilise le terme tirage au sort.

Clause d'ignorance : Fait de ne pas révéler à un sujet le médicament qu'il va recevoir dans un essai thérapeutique, et pour un médecin qui inclut un sujet dans un essai, de ne pas savoir (*avant l'inclusion*) quel traitement ce sujet va recevoir (*selon le groupe où il sera alloué*). Sinon, l'inclusion des sujets dans l'essai risque d'être influencée par la conviction intime du médecin de l'efficacité de l'un ou l'autre traitement réellement efficace. Le tirage au sort respecte la clause d'ignorance. (*définition du CNCI*).

B. Principe et intérêt de la randomisation

Procédé d'allocation au hasard (*random en anglais*) des sujets dans les groupes d'étude, elle permet:

- ✓ que chaque sujet ait la même chance, connue *a priori*, d'être assigné à un groupe donné
- ✓ de constituer des groupes comparables, en moyenne, sur tout sauf l'intervention, en particulier sur les facteurs pouvant avoir un effet sur le critère de jugement principal.
- ✓ donc d'éviter les biais de confusion
- ✓ d'éviter les biais de sélection, en ne permettant pas que l'affectation à un traitement soit le fruit d'un arrangement systématique et/ou soit liée au jugement de l'investigateur qui inclut les sujets.
- ✓ d'interpréter correctement les tests d'hypothèse; H₀ "pas de différence entre intervention et contrôle(s)" présuppose que seules les interventions différencient les groupes; on impute au traitement l'origine de la différence observée si H₀ est rejetée.
- ✓ et donc de respecter les contraintes éthiques de la méthode expérimentale en Sciences de la Vie.

C. Faire une randomisation

- **Méthode** : tirage au sort de l'allocation des sujets à un groupe ou à un autre par utilisation d'une table de nombres au hasard
- **Mode** :

- ✓ **Randomisation simple** : allocation aléatoire des sujets dans les groupes.
- ✓ **Randomisation stratifiée** : le but est d'équilibrer les effectifs de chaque groupe vis à vis de certaines caractéristiques et ainsi de forcer la comparabilité des groupes pour ces caractéristiques
 - on détermine des catégories de sujets, appelées strates, en fonction d'une ou de quelques caractéristiques (maximum 2 ou 3), connues pour modifier potentiellement l'effet du traitement - (ex: centre dans un essai clinique multicentrique, classes d'âge, service clinique ...). Au sein d'une strate, les individus sont répartis par tirage au sort entre les groupes de traitement.
 - exemple :
 - Stratifier **par centre** revient à faire une liste de randomisation indépendante et différente pour chaque centre. Par exemple, dans un essai avec deux groupes de traitement et un rapport 1:1 entre les groupes, la randomisation stratifiée par centre permet d'équi-répartir à l'inclusion les patients d'un même centre sur les deux groupes de traitement. La randomisation stratifiée par **centre et par sexe** permet d'équi-répartir à l'inclusion, sur les deux groupes de traitement, les patients du même sexe et du même centre. S'il y a n centres, il y aura $2n$ strates (n pour les hommes et n pour les femmes).
 - dans un essai clinique multicentrique, la randomisation **doit** être stratifiée systématiquement sur le centre (par exemple: si le rapport est 1:1, dans chaque centre on aura autant de patients recevant le traitement A que le traitement B) pour forcer la comparabilité a priori des groupes de traitement pour les facteurs pronostiques au sein de chaque centre. On neutralise théoriquement l'effet centre. Généralement on ajuste aussi sur le centre dans l'analyse.
 - Quand des analyses en sous-groupes sont prévues, la randomisation est stratifiée sur les variables qui définissent les sous-groupes (cf § objectif 9)
- ✓ **Randomisation par blocs**: le but est d'équilibrer les effectifs de chaque groupe au fur et à mesure de l'allocation des sujets. Un bloc est un sous-groupe de taille déterminée à l'avance (exemple 6) à l'intérieur duquel on a une allocation aléatoire des patients. Ex pour 2 traitements A et B et des blocs de 6: on a 3 sujets dans chacun des groupes A et B, ABAABB, BAAABB, ABABAB....Ainsi au bout de 6 sujets inclus, 3 sont alloués au traitement A et 3 au traitement B. L'investigateur ne doit pas connaître la taille du bloc afin de ne pas pouvoir prévoir le groupe du ou des derniers sujets du bloc. Pour ce type de randomisation, **la meilleure imprévisibilité est obtenue avec des blocs de taille variable définie aléatoirement avec permutation de séquence. La randomisation par blocs contribue à neutraliser l'effet cohorte.**

- **Equilibre de taille des groupes:**

- ✓ Randomisation équilibrée: les groupes doivent avoir la même taille
- ✓ Randomisation déséquilibrée: les groupes sont de tailles différentes; par exemple pour 1 patient randomisé dans le groupe A, 3 sont randomisés dans le groupe B.

- **Organisation:** la liste (séquence) de randomisation doit être élaborée dans un centre indépendant (procédure centralisée) avant toute inclusion de sujet dans l'étude et son accessibilité se fait à l'heure actuelle le plus souvent soit par un serveur vocal, soit par internet, avec traçabilité de l'inclusion des patients pour empêcher des retraits d'inclusion.

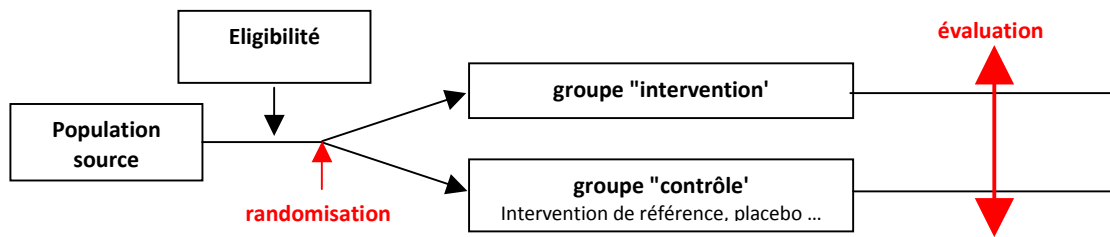
Moment de l'allocation du sujet: après confirmation de l'éligibilité +++ et le plus tard possible c'est à dire juste avant la première administration de l'intervention (pour éviter des exclusions post randomisation)

Cas particuliers

- ✓ Les essais en grappe (essais en cluster): on randomise les grappes (groupes homogènes quant à la question étudiée ex des classes d'école) et tous les sujets d'une même grappe (ex tous les élèves de la même classe) reçoivent la même intervention
- ✓ Essai croisé (ou en cross-over) : le patient est son propre témoin. On randomise la séquence de traitements.

Les détails concernant toute particularité de cette randomisation (stratification) doivent être précisés car il devra en être tenu compte dans l'analyse.

D. Validité d'une randomisation



- L'allocation du traitement ne doit pas pouvoir être prédite (respecter la clause d'ignorance)
- les modalités doivent être adaptées à l'objet et à la question
- La mise en œuvre doit être faite au moment opportun

E. Selon l'objet de l'étude:

La randomisation n'a lieu d'être **que dans les études interventionnelles comprenant plusieurs groupes soumis à comparaison.**

Dans les essais thérapeutiques comparatifs, la randomisation est la règle.

Dans les études où on veut comparer l'efficacité ou l'utilité d'une procédure de dépistage

Remarque : Dans certaines études le tirage au sort des patients (ou sondage) effectué pour obtenir un échantillon représentatif de la population à laquelle on veut extrapoler les résultats (ex : témoins dans une étude cas-témoins) est parfois appelé randomisation. Le terme de randomisation est utilisé alors à mauvais escient.

OBJECTIF 6 : DISCUTER LA COMPARABILITE DES GROUPES SOUMIS A LA COMPARAISON

1. LES QUESTIONS A SE POSER

- Les groupes analysés sont-ils comparables initialement en ce qui concerne les facteurs pronostiques ou de confusion connus ou suspectés? (un tableau présente-t-il les données permettant de s'assurer de cela ?)
- Les groupes analysés sont-ils comparables, sur ces mêmes facteurs, à la fin de l'étude ?
- Les perdus de vue, les sorties d'étude ou données manquantes sont-ils nombreux et déséquilibrés ?
- Dans un essai clinique, les groupes analysés ont-ils reçu la même prise en charge en dehors des interventions étudiées, au cours du suivi ?

2. OU CHERCHER

- Comparabilité initiale: dans le tableau donnant les résultats (nombre de sujets, fréquences et/ou moyennes) dans chaque groupe concernant les caractéristiques initiales des sujets étudiés, facteurs pronostiques majeurs ou de confusion.
- Suivi d'étude : section résultats et dans le **diagramme de flux** (flow chart) - ou son équivalent textuel - au début de la partie résultats, parfois à la fin de la partie méthode (perdus de vue, sorties d'étude).

3. RAPPELS THEORIQUES :

A. Suivi du nombre de sujets:

Il est important de connaître le nombre de sujets, éligibles (remplissant les critères d'inclusion), le nombre de sujets inclus et les raisons de non inclusion (refus, oubli, exclusion volontaire...).[cf objectif n°7]

B. Tableau de comparabilité des groupes (comparabilité initiale)

La comparabilité doit être clinique et non statistique car le critère de jugement principal sur lequel est fondée la conclusion de l'étude n'est pas un des facteurs pronostiques retenus pour évaluer la comparabilité. L'effectif n'a pas été calculé pour détecter une différence entre les groupes pour les facteurs pronostiques.

Quand l'étude est randomisée:

La randomisation des sujets entre les groupes garantit en théorie la comparabilité de ces groupes et prévient les biais de sélection.

Cependant, cette garantie n'est pas totale et la "comparabilité des groupes avant mise en œuvre de l'intervention" doit être vérifiée, en particulier sur les variables susceptibles d'intervenir sur les résultats (facteurs pronostiques).

S'il existe des différences dans les caractéristiques initiales (données de base), les auteurs doivent indiquer s'ils en ont tenu compte dans l'analyse (ajustement, analyse multifactorielle ou multivariée).

Quand l'étude n'est pas randomisée:

Les différences observées entre les groupes comparés peuvent s'expliquer par de nombreux facteurs autres que celui que l'on étudie.

Il est normal dans de telles études que les groupes ne soient pas comparables pour certaines variables. Pas d'imputation causale directe (voir faisceau d'arguments).

Il faut donc s'assurer que ces variables, lorsqu'elles interfèrent avec les expositions qui font l'objet de l'étude et avec la maladie étudiée, ont été prises en compte dans l'analyse : ajustement, analyse multifactorielle).

C. Diagramme de flux (maintien de la comparabilité initiale)

Il renseigne sur :

- le type d'analyse effectuée dans le cas d'un essai thérapeutique (analyse en intention de traiter) (cf objectif 9) : nombre de sujets exclus de l'analyse et motifs d'exclusion de l'analyse
- les sorties d'étude (nombre et motifs)

Si les motifs d'exclusion de l'analyse ou de sortie d'étude sont liés à l'intervention et/ou au phénomène mesuré, on a un biais de sélection. Ce biais est faible si ces exclusions ou sorties d'étude sont peu nombreuses (10% est la limite communément admise) et équilibrées entre les groupes. Il est élevé si elles dépassent 20%, et l'est d'autant plus qu'elles sont déséquilibrées.

- **Les traitements concomitants et différences de prise en charge pendant le suivi, le cas échéant**

Toute prise en charge ne faisant pas partie de l'intervention et concomitante à l'intervention, susceptible d'avoir un effet sur le phénomène étudié engendre un biais (cas des traitements concomitants).

Le phénomène observé ne sera pas dû uniquement à l'intervention ou au facteur étudié. Ce biais est d'autant plus important que la prise en charge est différente entre les groupes comparés.

OBJECTIF 7 : DISCUTER L'EVOLUTION DES EFFECTIFS ETUDIES ET LEUR COHERENCE DANS LA TOTALITE DE L'ARTICLE ; SAVOIR SI LE CALCUL DU NOMBRE DE SUJETS NECESSAIRES A ETE EFFECTUE A PRIORI

1. LES QUESTIONS A SE POSER

- a. Le nombre de sujets : éligibles, inclus, suivis ... peut-il être suivi dans l'article ? Le diagramme de flux et les tableaux de résultats permettent-ils de suivre les effectifs ?
- b. Un calcul du nombre de sujets nécessaire a-t-il été fait ?
- c. Ce calcul porte-t-il sur le critère de jugement principal ? Si non, le choix du critère de jugement est-il justifié ?
- d. Dans une étude comparative, la différence cliniquement "intéressante" (Δ) de valeur du critère de jugement principal entre les groupes est-elle fournie et argumentée ? Le risque de première espèce (α) est-il fourni ainsi que la puissance ($1-\beta$) ? Sont-ils satisfaisants ? Si le critère de jugement est qualitatif ou censuré, le taux dans le groupe contrôle est-il précisé ?
- e. Dans une étude non comparative dont l'objectif est l'estimation de paramètres (voir le critère de jugement principal), il permet de garantir une précision de cette estimation (ex : estimation de prévalence d'une maladie dans une enquête transversale, estimation des Se, Sp, VPP d'un examen dans le cadre d'une étude diagnostique). Le taux attendu, la précision et le risque α sont-ils donnés ?
- f. Le nombre de sujets sélectionnés tient compte du % potentiel de perdus de vue.
- g. Le nombre de sujets analysés est-il supérieur ou égal au nombre de sujets nécessaires

2. OU CHERCHER

Dans la partie méthode, pour le calcul

Dans la partie résultats, pour le suivi des effectifs et le diagramme de flux.

3. RAPPELS THEORIQUES :

Importance du calcul du nombre de sujets nécessaire (NSN)

- ✓ Le NSN est calculé en fonction du critère de jugement principal (CJP).
- ✓ Il permet de garantir à l'étude une puissance pour détecter une différence cliniquement intéressante du CJP entre les groupes compte tenu de la dispersion du CJP et des éventuelles sorties d'étude.
- ✓ La puissance doit au moins être égale à 80%
- ✓ Si le résultat n'est pas significatif, cela peut être dû à un manque de puissance, un biais ou à une différence plus petite que celle qu'on voulait détecter.
- ✓ Dans une étude non comparative, il permet de garantir une précision de l'estimation du CJP.

La méthode de calcul du nombre de sujets doit être indiquée avec la différence que l'on veut mettre en évidence entre les groupes comparés, (ou la précision des indicateurs évalués) et les risques d'erreur acceptés (en général $\alpha = 5\%$ et β entre 5 et 20%).

La "plausibilité" de la différence "intéressante" repose sur une argumentation.

« Le CONSORT statement » propose un diagramme-type (flow chart) permettant de s'assurer que le lecteur retrouve clairement le nombre de sujets concernés à chaque étape - recrutement des patients, inclusion, randomisation, suivi, analyse de l'étude - ainsi que les motifs d'exclusion/non suivi de l'étude.

OBJECTIF 8 : S'ASSURER QUE LA METHODE EMPLOYEE EST COHERENTE AVEC LE PROJET DU TRAVAIL, QUE LA METHODOLOGIE EST EFFECTIVEMENT SUSCEPTIBLE D'APPORTER UNE REPONSE A LA QUESTION POSEE DANS L'INTRODUCTION.

1. LES QUESTIONS A SE POSER

Le schéma d'étude est-il cohérent avec l'objet ? Le schéma d'étude est-il cohérent avec la question étudiée? (Pour la justification voir articles vus en enseignements dirigés).

A.

2. OU CHERCHER

En général cité dans le titre de l'article puis précisé dans le chapitre méthode. Parfois en fin d'introduction avec le (les) objectif(s) de l'étude.

3. RAPPELS THEORIQUES :

A. Définition du Schéma d'étude ou Type d'étude.

Le schéma d'étude est la méthode stratégique mise en œuvre pour répondre à la question posée. C'est sur lui, en adéquation avec la question posée, que repose toute la valeur de l'article.

Selon qu'il s'agit d'un problème diagnostique, thérapeutique, pronostique ou épidémiologique, le schéma d'étude varie.

[voir addendum n°2 pour le détail des schémas d'étude]

QUESTION		TYPE d'ENQUETE / ETUDE les plus adaptées
Epidémiologie descriptive	Incidence	Cohorte
	Prévalence	Transversale
Epidémiologie analytique (causalité)	Phénomène non contrôlable fréquent	Cohorte exposés / non exposés
	Phénomène rare	Cas - Témoins
Traitement	Efficacité	Essai contrôlé randomisé
Traitement	Sécurité	Essai contrôlé randomisé Cohorte exposés / non exposés
action de dépistage	Efficacité	Essai contrôlé randomisé Cohorte exposés / non exposés
Procédure diagnostique	Fiabilité, (reproductibilité/variabilité)	Transversal comparatif avec répétition de mesure
	Validité (sensibilité/spécificité)	Transversal comparatif avec référence (Gold standard, étalon or) (sujet propre témoin)
Pronostic	Maladie fréquente	Cohorte exposés / non exposés
	Maladie rare	Cas - Témoins

ANAES / Service Recommandations Professionnelles / Janvier 2000

B. Schéma d'étude par Objet d'article

a. Le schéma d'étude des enquêtes épidémiologiques

1. Enquêtes descriptives:

but : décrire des indicateurs mesurant l'importance de la maladie : incidence, prévalence, mortalité.

Schémas : transversal pour la prévalence ou longitudinal (pour l'incidence ou la mortalité): cohorte. Les registres sont des enquêtes à recueil permanent, prospectif et dont le but est l'exhaustivité et l'exactitude, utilisés pour obtenir des incidences.

Sujets : soit échantillon représentatif d'une population définie (sondage), soit population entière.

2. Enquêtes analytiques

but : mettre en évidence des facteurs de risque de maladies

Schéma : cohorte, cohorte exposés-non exposés, cas-témoins, registre (recueil permanent)

Sujets : 2 groupes (exposés / non exposés: cohorte exposés / non exposés ou malades / non malades: cas témoins). Dans les enquêtes de cohorte où on recherche l'association entre plusieurs facteurs d'exposition et une maladie, on recueille les informations concernant les facteurs d'exposition chez tous les sujets (on ne distingue pas de groupes).

b. Le schéma d'étude des études diagnostiques

Au sein des études diagnostiques, deux grands types:

- (1) évaluation de la fiabilité d'un examen ou procédure
- (2) évaluation de la validité d'un examen ou procédure

Evaluation de la fiabilité d'un examen ou procédure

Etude de concordance dans laquelle on fait évaluer le même examen par plusieurs juges et on étudie la concordance des résultats entre les différentes évaluations. C'est une étape préalable à l'examen de la validité.

Evaluation de la validité d'un examen ou procédure

On peut soit :

- ✓ vouloir estimer la sensibilité et la spécificité d'un nouveau test,
- ✓ déterminer un seuil à partir duquel d'autres examens ou un traitement seront mis en oeuvre,
- ✓ comparer un test diagnostique à un autre (ou une séquence à une autre).

Point clé à vérifier : le nouveau test et le gold standard doivent mesurer le même état du sujet, c'est pourquoi le meilleur schéma est une étude expérimentale **transversale**.

c. Le schéma d'étude des études d'estimation d'un pronostic

Ce sont des études à visée analytique.

On distingue deux types d'étude pronostique : prospectives et rétrospectives.

Les études prospectives consistent à suivre un groupe de malades (cohorte) pendant une durée définie, d'enregistrer les facteurs que l'on veut étudier (facteurs potentiellement pronostiques) et de noter les événements (complications, décès) au fur et à mesure de leur apparition.

Les études rétrospectives :

- Ce sont le plus souvent des cohortes : elles consistent à reconstituer une cohorte *a posteriori* à partir d'une file active et à aller chercher dans les dossiers pour une période antérieure les éléments (facteurs pronostiques et événements) nécessaires à l'étude. Elles pâtissent souvent de la qualité des dossiers ...
- Plus rarement ce sont des études cas-témoins

d. Le schéma d'étude des études d'évaluation d'un traitement

C'est l'essai thérapeutique contrôlé randomisé en double aveugle. cf tableau §3.A . Voir glossaire (expressions comportant le terme *essai*)

acronyme: **ECHRAMS**

Essai : Essai clinique thérapeutique, de prévention,

Contrôlé contre X (citer traitement du groupe contrôle),

Hypothèse testée : d'efficacité, (OU de non infériorité, -OU d'équivalence)

Randomisé,

Aveugle? en double aveugle, (OU en ouvert OU en simple aveugle),

Multicentrique, (OU monocentrique).

Schéma : à X (2, 3) bras parallèles, (OU croisé = en cross-over),

e. Le schéma d'étude des études d'évaluation d'une action de dépistage **d'une maladie (application aux cancers)**

cf tableau §3.A

Attention : l'évaluation d'un **test de dépistage** utilise les mêmes méthodes que l'évaluation diagnostique (seules les populations changent).

1. Evaluation d'une **procédure de dépistage** :

- Objectif : démontrer l'efficacité (et/ou la qualité) de la procédure de dépistage
- Design optimal :
 - o essai contrôlé randomisé
 - o 2 bras : 1 bras effectuant le dépistage, un bras n'effectuant aucun dépistage
- Biais spécifiques : avance au diagnostic (lead-time), biais de surdiagnostic (famille des biais de classement)
- Critère de jugement principal :
 - o Si l'objectif est l'efficacité : le critère est la mortalité spécifique (car permet de s'affranchir du biais d'avance au diagnostic : ex dépistage des cancers), ou l'incidence du cancer si l'objectif est de dépister des lésions pré-cancéreuses
 - o si l'objectif est l'évaluation de la qualité : les critères de jugement sont la sensibilité, spécificité et valeurs prédictives de la *procédure* (test+bilan diagnostique)
 - o Sensibilité de la procédure : impose l'existence d'une structure qui enregistre de façon exhaustive les cas de maladie (cancers de l'intervalle)
 - o Spécificité de la procédure: impose que l'on dispose d'une organisation qui permette de suivre jusqu'au terme du bilan diagnostique tous les sujets ayant un test de dépistage « positif »
- Analyse en intention de dépister
- Si suivi long : problème de la contamination des groupes

2. Evaluation d'un **programme de dépistage** :

- Objectif : vérifier que, lors de la généralisation du dépistage à une population ou un sous-groupe de population (dépistage organisé), le programme de dépistage respecte les conditions dans lesquelles on peut attendre des résultats similaires à ceux des essais d'efficacité.
- Evaluation le plus souvent selon un cadre réglementaire (ex : dépistages des cancers).
- Critères d'évaluation calculés sur le sous-groupe de population concerné par le dépistage (participants ou non)
- Critères d'évaluation principaux : participation, qualité, efficacité du programme, respect des délais de prise en charge, parfois évaluation des conséquences psychologiques et sociales du dépistage.
- Dans le cas particulier des programmes de dépistage des cancers, les critères d'efficacité sont d'abord des critères *intermédiaires* (le calcul du taux de mortalité demandant le plus souvent un long délai) portant essentiellement sur la taille et le degré d'envahissement (facteurs pronostiques majeurs) de la maladie.

C. Exemples:

Les exemples suivants (questions de recherche et les plans expérimentaux appropriés) permettent de comprendre les réponses à donner à cette question.

1) **Objet de l'article**: Evaluation de l'efficacité thérapeutique d'une intervention médicale

- Exemple d'objectif : Mesurer (Evaluer) l'efficacité d'un médicament antivitaminique K pour prévenir la survenue d'accident thrombo-embolique chez des sujets de plus de 65 ans, en arythmie cardiaque par fibrillation atriale.
- Hypothèse : antivitaminique K réduit le risque de survenue d'accident thrombo-embolique chez des sujets de plus de 65 ans, en arythmie cardiaque par fibrillation atriale.
- Type d'étude: essai thérapeutique d'efficacité multicentrique, contrôlé, randomisé, en double aveugle, à deux groupes parallèles : Antivitaminique K versus placebo,

2) **Objet de l'article** : Evaluation d'une procédure de dépistage

- Exemple d'objectif : Déterminer si la prescription systématique d'un test PSA à une population masculine de 50 à 75 ans permet de réduire la mortalité par cancer de la prostate.
- Hypothèse : la prescription systématique d'un test PSA à une population masculine de 50 à 75 ans réduit la mortalité par cancer de la prostate.
- Type d'étude: essai contrôlé randomisé prospectif à deux bras parallèles avec prescription d'un dépistage systématique versus non prescription de dépistage

3) **Objet de l'article : Evaluation d'une intervention**

- Exemple d'objectif: La mise en place de séances d'information sur les risques liés au tabagisme dans les écoles primaires permet-elle de diminuer le nombre de fumeurs chez les adolescents ?
- Hypothèse : Une information sur les risques liés au tabagisme dans les écoles primaires diminue le nombre de fumeurs chez les adolescents.
- Type d'étude : essai randomisé entre les diverses écoles d'une même ville : essai en cluster contrôlé randomisé, séances d'information versus pas de séance

4) **Objet de l'article : Enquête épidémiologique**

Descriptive

- Exemple d'objectif : Déterminer la prévalence des infections nosocomiales dans les établissements de courts séjours publics et privés français.
- Hypothèse : la prévalence des infections nosocomiales dans les établissements de courts séjours publics et privés français est de x%.
- Type d'étude: étude épidémiologique d'observation transversale.

Recherche de facteurs de risque

- Exemple d'objectif : Déterminer le risque de sclérose en plaques lié à la vaccination contre l'hépatite B.
- Hypothèse : le vaccin contre l'hépatite B est associé à un risque accru de sclérose en plaque.
- Type d'étude: étude épidémiologique d'observation rétrospective (événement rare) à visée analytique de type cas-témoin

5) **Objet de l'article : Evaluation d'une procédure diagnostique**

- Exemple d'objectif : Evaluer la valeur diagnostique de la scintigraphie pulmonaire ventilation/perfusion, chez un patient avec suspicion d'embolie pulmonaire
- Hypothèse : la scintigraphie pulmonaire ventilation/perfusion permet de détecter une embolie pulmonaire.
- Type d'étude : étude transversale avec sujet propre témoin

6) **Objet de l'article : Estimation d'un pronostic : identifier des critères pronostiques**

- Exemple d'objectif : Identifier les facteurs qui augmentent le risque de séquelles chez les enfants atteints de méningite bactérienne
- Hypothèse : Les facteurs X, Y, Z augmentent le risques de séquelles chez les enfants atteints de méningite bactérienne-
- Type d'étude : étude observationnelle de cohorte prospective à visée analytique.

OBJECTIF 9 : VÉRIFIER QUE LES ANALYSES STATISTIQUES (EN FONCTION DE NOTIONS ÉLÉMENTAIRES) SONT COHÉRENTES AVEC LE PROJET DU TRAVAIL ; CONNAÎTRE LES LIMITES DE L'ANALYSE PAR SOUS GROUPE ; CONNAÎTRE LA NOTION DE PERDUS DE VUE.

1. LES QUESTIONS A SE POSER

Certains aspects sont un peu complexes ; nous nous contenterons de lister quelques points-clés simples à vérifier :

- a. La stratégie d'analyse est-elle cohérente avec le projet du travail ?
 - i. Si l'étude est un essai clinique d'efficacité (de supériorité ou de différence), l'analyse est-elle réalisée en intention de traiter ?
 - ii. Si l'étude est un essai clinique de non infériorité, deux analyses sont-elles réalisées, l'une en intention de traiter, l'autre en per protocole, et leurs résultats ont-ils été confrontés ?
 - iii. Si l'étude est observationnelle à visée analytique, une analyse multivariée avec ajustement sur les facteurs de confusion ou sur les facteurs pronostiques a-t-elle été effectuée ?
- b. Les tests choisis sont-ils adaptés au type de données analysées ?
- c. Les intervalles de confiance des résultats sont-ils précisés ?
- d. Si des analyses en sous-groupes ont été effectuées, étaient-elles prévues dans le protocole, argumentées, avec un but précisé (exploratoire, OU confirmatoire)? La méthode utilisée permet-elle de fournir une conclusion valide pour ces analyses ? (La randomisation a-t-elle été stratifiée ? L'interaction a-t-elle été testée ? A-t-on corrigé le risque alpha ? L'interprétation repose-t-elle sur l'interaction ? Le calcul du NSN a-t-il tenu compte des analyses en sous-groupes ?)
- e. S'il y a des perdus de vue, engendrent-ils une perte de puissance importante ? Engendrent-ils un biais majeur ?
- f. S'il y a une (des) analyses intermédiaires,
 - i. Ont-elles été prévues dans le protocole ?
 - ii. A-t-on tenu compte de ces analyses intermédiaires dans le calcul du NSN ?
 - iii. A-t-on précisé un seuil de signification pour chacune d'elles pour fonder les conclusions statistiques ?

Pour les notions de causalité se reporter au point D et à l'objectif 17

2. OU CHERCHER

Stratégie d'analyse et analyses statistiques : partie méthode: analyse, partie résultats, diagramme de flux (stratégie d'analyse)

Pour les analyses en sous-groupes : planification et argumentation : partie introduction

Planification: partie méthode : randomisation et analyses

Interprétation : partie résultats et discussion.

Pour les perdus de vue : partie méthode (NSN) et partie résultats et discussion (biais, puissance).

Pour les analyses intermédiaires : parties méthode et résultats

3. RAPPELS THÉORIQUES

A. Intention de traiter

Définition: Analyse en intention de traiter = Méthode qui consiste à analyser les données de tout patient inclus, et ce dans le « bras » (groupe de tirage au sort) dans lequel il a été randomisé au début de l'étude.

Cette analyse permet:

- De ne pas détruire la randomisation
- D'analyser les données de façon pragmatique, comme dans la réalité clinique (le traitement pouvant être non actif chez certains patients et donc changé par le praticien). On compare donc les deux interventions (intervention à tester et intervention contrôle) sur leurs effets et ce dans les conditions les plus proches de "la pratique courante".
- Donc d'analyser les sujets randomisés dans leur groupe de randomisation:
 - ✓ Quelle que soit leur observance au traitement
 - ✓ Quel que soit le traitement réellement reçu
 - ✓ Quel que soit l'éventuel retrait du patient de l'étude ou l'éventuelle déviation au protocole.

Vérification

- Les auteurs auront dans la plupart des cas noté, en partie méthodologie - analyse -, que leur analyse a été faite "en intention de traiter".
- La vérification s'appuie sur le tableau du flow-chart ainsi que sur les tableaux de résultats en comparant le nombre de patients randomisé et le nombre de patients analysés dans chaque groupe de traitement.

Cette méthode se distingue de

- "l'analyse per protocole" où seuls les sujets ayant suivi jusqu'au bout le protocole sont analysés (des sujets seront exclus de l'analyse sur la base d'informations observées après leur randomisation : non observants, changement de posologie ou arrêt de traitement par le médecin, perdus de vue, permutation de traitement...)
- et de « l'analyse en traitement réellement reçu » où les sujets sont analysés selon le traitement effectivement reçu, même s'il ne s'agit pas du traitement qui leur avait été attribué initialement par la randomisation.

Comparaison des résultats et conclusions issus des analyses per protocole ou en traitement réellement reçu et en ITT.

En général

- Différence estimée avec l'analyse per protocole plus grande qu'avec ITT.
- Puissance moins élevée avec analyse per protocole qu'avec ITT
- Biais de sélection potentiels plus élevés avec analyse per protocole qu'avec ITT.
- Tendance de l'ITT à être conservatrice (ne pas mettre en évidence une différence qui existe vraiment).

Remarque : Moins il y a d'écart au protocole, de perdus de vue, d'arrêts prématurés des traitements, de données manquantes, plus les résultats des deux types d'analyses sont proches. En l'absence d'écart au protocole, de sorties d'étude, d'arrêts prématurés des traitements, de données manquantes, les résultats des deux analyses sont identiques.

B. Les tests choisis sont-ils adaptés à la question et au type de données analysées

On vérifie que les tests classiques pour données qualitatives ou quantitatives, décrits dans la partie méthode, et référencés, ont été utilisés, de même que les tests spécifiques en cas d'écart aux conditions de validité des tests classiques (ex : faible nombre de sujets (<30), distribution non gaussienne), de séries appariées, de type d'étude particulier.

Les tests doivent être appropriés aux données, et il est donc indispensable de vérifier l'adéquation du test statistique avec :

- Le type de la variable,
- La distribution des variables. Les variables non gaussiennes doivent le plus souvent être analysées avec des tests non paramétriques, (c'est le cas des petits effectifs),
- Le type d'expérience : groupes distincts (parallèles) ou mesures répétées chez un même sujet.

(Le détail des tests à utiliser en fonction de toutes ces données est joint en annexe 1. Une liste de références bibliographiques adaptées est jointe en annexe 2).

Si l'objectif de l'étude était une comparaison de groupes, vérifier qu'un test statistique a bien été effectué pour comparer ces groupes.

Si l'étude est observationnelle avec une recherche de facteurs de risque ou de facteurs pronostiques, vérifier qu'un ajustement sur les facteurs de confusion ou les autres facteurs pronostiques a été réalisé.

C. Les intervalles de confiance des résultats sont-ils précisés ?

L'intervalle de confiance d'un résultat indique les limites à l'intérieur desquelles la "vraie" valeur estimée (différence entre traitements, risque relatif, sensibilité etc...) est susceptible de se trouver. Qu'il s'agisse de pourcentage, de moyenne, de risque, il est indispensable d'assortir les résultats de leur intervalle de confiance (en général à 95%, soit $\alpha = 5\%$), ce qui permet de juger de la précision des résultats et des valeurs auxquelles on peut s'attendre lorsque l'on appliquera les résultats dans la pratique.

D. Analyses en sous-groupes

1. Il existe trois cas

- Analyses en sous-groupes planifiées (prévues dans le protocole)
 - A visée de confirmation
 - A visée exploratoire
- Analyses post-hoc (non planifiées, non prévues dans le protocole)

Si elles n'ont pas été planifiées comme analyses à visée de confirmation, les analyses en sous-groupe sont toujours exploratoires. Aucune conclusion clinique ne peut être fondée sur les résultats des ces analyses non planifiées ou à visée exploratoire.

2. Conditions nécessaires pour fonder une conclusion clinique. Les analyses en sous-groupe doivent être

- planifiées comme analyse à visée de confirmation,
- justifiées (cf arguments § E.5),
- limitées en nombre
- la randomisation doit avoir été stratifiée sur la variable qui définit les sous-groupes (*exemple genre*).
- l'essai doit avoir la puissance suffisante pour détecter les interactions (remarque : généralement le NSN n'est pas calculé pour garantir une puissance pour détecter une certaine interaction pour des raisons de faisabilité et de coût).
- l'interaction doit avoir été testée.
 - **Définition interaction** (glossaire) : Mesure dans laquelle l'effet d'un facteur est modifié en fonction de l'action d'un ou de plusieurs autres facteurs. *Ex : interaction du genre sur l'effet d'un traitement*
 - **Interaction quantitative** : le traitement est bénéfique (ou délétère) dans tous les sous-groupes définis par une variable mais a une taille d'effet différente. *EX: effet bénéfique du traitement A deux fois plus important chez les femmes que chez les hommes.*
 - **Interaction qualitative** : le traitement est bénéfique dans un sous-groupe mais est délétère dans un autre. *Ex: effet du traitement B bénéfique chez les femmes mais délétère chez les hommes.*

3. Conclusions d'une analyse en sous-groupe

- **Si l'essai est concluant** (différence significative quand on analyse tous les sujets globalement), **si l'interaction est très significative, et si elle est justifiée par un argument fort** (voir ci-dessous), **les résultats des analyses en sous-groupes peuvent être pris en compte dans la conclusion clinique.**)
On peut alors conclure que l'effet du traitement est différent entre les sous-groupes : *exemple - traitement efficace chez les femmes mais délétère chez les hommes.*
- **Attention** aux interactions significatives par le fait du hasard.
- **Si l'essai est non concluant** (pas de différence significative quand on analyse tous les sujets), **les analyses en sous-groupes devraient être exploratoires.**
- **En règle générale, les résultats d'une analyse en sous-groupe nécessitent d'être confirmés dans une autre étude.**

4. Arguments justifiant les analyses en sous-groupe : ils peuvent être de nature diverse

- Exemples : Variation génétique
 - *Exemple en évaluation thérapeutique : Dans le cancer du colon, la réponse à la chimiothérapie dépend de l'expression de gènes. La conservation de l'hétérozygotie au niveau des loci des chromosomes 17p et 18q est un facteur prédictif de l'effet d'une chimiothérapie avec du fluorouracile. Le traitement augmente la survie chez les patients n'ayant pas perdu l'hétérozygotie. En revanche, il ne présente pas de bénéfice chez les patients ayant complètement perdu l'hétérozygotie.*
 - *Exemple en recherche épidémiologique: Infection au VIH et récepteur CCR5 des lymphocytes T CD4+. Le corécepteur CCR5 est indispensable à l'entrée du VIH dans les lymphocytes T CD4+.*

Certains sujets homozygotes pour la délétion de la séquence génétique qui rend CCR5 non fonctionnel sont résistants à l'infection par le VIH.

E. Analyses intermédiaires

- **Définition (glossaire)** : Analyse effectuée avant l'inclusion de tous les sujets prévus. Elle est réalisée le plus souvent lorsque l'étude est longue. **Elle doit être prévue dans le protocole (argumentée), et le nombre de sujets nécessaires prend en compte le nombre d'analyses intermédiaires qui sont prévues. Pour chaque analyse intermédiaire, un seuil de signification doit être déterminé (plusieurs méthodes sont possibles).**
- Arguments pour justifier une analyse intermédiaire:
 - Les hypothèses de taille d'effet ne sont pas très précises.
 - i. Dans le cas où on prévoit un bénéfice net précoce : il n'est pas éthique de priver des patients d'un traitement qui s'avérerait plus efficace que prévu. le principe d'équipoise entre les groupes de traitements n'est plus respecté, autrement dit, il n'y a plus de réelle incertitude à l'intérieur de la communauté scientifique quant au meilleur traitement. : arrêt de l'essai pour efficacité.
 - ii. De la même façon, il n'est pas éthique de soumettre des patients à un traitement qui s'avérerait moins efficace que prévu (arrêt de l'essai pour futilité)
 - Les effets secondaires attendus sont potentiellement graves pouvant remettre en question la balance bénéfice-risque : il n'est pas éthique de poursuivre un essai dans ces conditions : arrêt de l'essai pour sécurité.
- Pour chaque analyse intermédiaire, le seuil de signification doit avoir été corrigé pour garantir un risque α global. Ex : si α global fixé à 5%, $\alpha_{int} \ll 5\%$.
- Dans tous les cas, les règles d'arrêt ou de poursuite de l'essai doivent être définies explicitement a priori.

F. Perdus de vue

- **Définition (glossaire)** : Patient qui n'est pas suivi sur la totalité de la période prévue par le protocole d'un essai ou d'une étude épidémiologique. On ne sait pas si le patient a guéri, s'il a eu une complication ou des effets secondaires, et pourquoi il n'est pas revenu.
- **Conséquences des perdus de vue**
 - Perte de précision et/ou perte de puissance dans une comparaison
 - Biais de sélection si les perdus de vue sont différents (pour les facteurs pronostiques) des patients restés dans l'étude. Biais négligeable si taux < 10%, acceptable si $10\% \leq$ taux < 20%, important si taux $\geq 20\%$ (taux empiriques). Le biais est d'autant plus important que les taux de perdus de vue sont déséquilibrés entre les groupes comparés. Ex: 30% vs 10%

OBJECTIF 10 : VERIFIER LE RESPECT DES REGLES D'ETHIQUE

1. LES QUESTIONS A SE POSER

- Le protocole de l'étude a-t-il été soumis à un comité d'éthique indépendant avant l'inclusion du premier patient ? A-t-il reçu l'approbation du comité?
- Le consentement, libre et éclairé, de chaque patient a-t-il été recueilli ?
- Les données sont-elles anonymes ? Sinon ont-elles reçu une habilitation de stockage?

Les règles de confidentialité des données recueillies ont-elles été énoncées et respectées ?

- Si, dans un essai clinique, on utilise un placebo, cette utilisation est-elle justifiée ?

2. OU CHERCHER

Dans la partie méthode

Pour l'utilisation du placebo, parties introduction, méthodes, discussion.

3. RAPPELS THEORIQUES :

A. Les trois principes fondamentaux

Ils découlent de la **déclaration d'Helsinki**, adoptée au niveau international et donnant des recommandations concernant tous les participants à une recherche biomédicale et **des bonnes pratiques cliniques** (conférence internationale d'harmonisation (ICH : international conference of harmonization)).

Respect de la personne humaine:

Toute recherche implique une information juste, complète et accessible au sujet quant aux finalités et aux modalités de l'étude, ainsi qu'aux risques auxquels le sujet pourrait être exposé.

Le sujet doit donner son consentement libre et éclairé à sa participation à l'étude

Principe d'utilité

Toute recherche doit amener un "bien".

La balance bénéfico-risque doit être en faveur du bénéfice pour le participant.

Principe de justice

Les sujets, égaux en dignité et en droits, doivent être traités de façon équitable.

Les populations vulnérables ont besoin d'une protection spéciale. 'La recherche médicale impliquant une population ou une communauté défavorisée ou vulnérable se justifie uniquement si la recherche répond aux besoins et priorités sanitaires de cette population ou communauté et si, selon toute vraisemblance, les résultats de la recherche seront bénéfiques à cette population ou communauté (déclaration d'Helsinki).

B. Conséquences

1. Toute recherche bio-médicale sur un individu, son dossier médical ou des produits issus du corps humain (produits sanguins, tissus...) est soumise à une réglementation stricte (*en France, loi du 20 décembre 1988 modifiée en 2004, dite loi "Huriet-Sérusclat", et en mars 2012 (dite loi "Jardé") lois de bioéthique et loi "informatique et libertés"*) et donc à l'approbation de comités et d'instances légales.
2. Le projet doit avoir été soumis par l'investigateur à un **Comité d'Ethique Indépendant** (*CPP ou Comité de Protection des Personnes, en France, ethical committee, ailleurs*). Parfois l'accord du comité institutionnel (*Institutional Review Board, IRB*) est mentionné. C'est une instance locale qui ne garantit pas l'indépendance par rapport aux investigateurs et/ou au promoteur. Dans le cadre d'un projet international: soumission à un comité par pays. L'approbation doit avoir été donnée avant toute inclusion dans l'étude.
3. Corollaire:
 - ✓ Dans le cas où l'accord du Comité d'éthique n'a pas été obtenu avant le début de l'étude, les auteurs doivent le noter et en donner les raisons.
 - ✓ Dans le cas où la recherche ne nécessiterait pas l'accord du Comité, les auteurs doivent en spécifier les raisons et spécifier comment ils se sont assurés que leur travail sortait bien de ce cadre.
 - ✓ Toute modification du protocole nécessite un accord complémentaire du (des) comité(s)

4. Tout patient doit donner son **consentement libre et éclairé** à sa participation à l'étude. En cas contraire, le patient ne peut être inclus. Sa non inclusion doit être notée dans le tableau de flux. Un patient peut se retirer à tout moment de l'étude.
5. Le promoteur de l'étude doit avoir souscrit une assurance couvrant les risques pour les participants, le cas échéant.
6. Toute collecte automatisée de données permettant l'identification, même indirecte, des participants, doit avoir été soumise au Comité d'Ethique (*pour la France, à la CNIL, Commission Informatique et Libertés*) et reçu un avis favorable préalablement à toute inclusion.
7. En cas de recherche impliquant la personne humaine de type interventionnel, l'avis de l'autorité compétente (*en France AFSSAPS, Agence française de sécurité sanitaire des produits de santé*) est obligatoire.

Remarque: Ces trois derniers points ne sont généralement pas mentionnés dans les articles publiés dans les revues internationales

C. Conditions d'utilisation d'un placebo dans les essais cliniques

○ Déclaration d'Helsinki

Les bénéfices, risques, dangers, efficacité d'une nouvelle intervention doivent être comparés à ceux de la meilleure intervention actuelle reconnue (preuve scientifique établie de son efficacité), sauf dans les cas suivants:

- i. L'utilisation d'un placebo ou l'absence de traitement est acceptable dans les études où aucune intervention reconnue n'existe
 - ii. Quand pour des raisons irréfutables et scientifiquement valables, l'utilisation d'un placebo est justifiée pour évaluer l'efficacité et la tolérance d'une intervention, et les patients qui reçoivent le placebo
 - iii. d ne seront pas exposés à un risque/danger grave ou irréversible.
- Conséquences
 - i. Comité d'éthique juge si, au vu des connaissances du moment, l'utilisation d'un placebo est éthiquement acceptable => rechercher si l'étude a été approuvée par un comité d'éthique
 - ii. Rechercher si consentement libre et éclairé a été signé par les patients

OBJECTIF 11°: ANALYSER LA PRESENTATION, LA PRECISION ET LA LISIBILITE DES TABLEAUX ET DES FIGURES, LEUR COHERENCE AVEC LE TEXTE ET LEUR UTILITE.

1. LES QUESTIONS A SE POSER

- A. Quels sont les critères de qualité d'un tableau ?
- B. Quels sont les critères de qualité d'une figure ?
- C. Où chercher les informations ?

2. OU CHERCHER

Tout ce qui figure dans un tableau ou une figure concerne généralement les résultats de l'étude. C'est donc dans la section « Résultats » que se situent les tableaux et les figures, ainsi que le texte s'y rapportant.

3. RAPPELS THEORIQUES :

1. TABLEAUX

Les tableaux sont indiqués pour présenter soit des données répétitives sous forme synthétique soit des données utiles pour vérifier des résultats importants. Les données répétitives sont par exemple les données servant à vérifier la comparabilité des groupes dans un essai randomisé. Ils servent à gagner de la place et à gagner en clarté. Les critères de qualité d'un tableau à vérifier concernent les points suivants :

1. Sa taille doit être raisonnable. Un tableau trop grand présentant des types d'informations trop différentes est difficile à lire.
2. Il doit comporter un titre informatif situé au dessus du tableau et doit indiquer le contenu du tableau et la population concernée (individus, lieu, temps).
3. Les variables (ou facteurs) qu'il décrit doivent être identifiées clairement (les abréviations doivent être définies dans une note de bas de tableau), avec leur unité pour les variables quantitatives. Les effectifs de sujets sur lesquels sont donnés les résultats doivent être indiqués systématiquement (et pas seulement les pourcentages).
4. Les nombres figurant dans les cellules du tableau représentent les résultats : ceux-ci doivent être présentés de façon identique et regroupés pour toutes les variables du même type (qualitative versus quantitative)
5. Les indices de dispersion doivent être reportés. Il n'est pas informatif de donner une moyenne sans son écart type, ou une médiane sans les quartiles ou les valeurs extrêmes. Lorsqu'on donne les résultats d'une estimation telle que prévalence, incidence, risque relatif, odds ratio, sensibilité..., l'intervalle de confiance et son niveau de confiance doivent être donnés (*ex : sensibilité 97%; intervalle de confiance à 95% = [87,9 ; 99,6]*).
6. Les valeurs statistiquement significatives doivent être clairement identifiables.
7. Il doit pouvoir se lire de façon autonome (auto-suffisant : il n'est pas nécessaire de s'aider du texte pour en comprendre le sens).
8. Il doit être référencé dans le texte.
9. Les informations qu'il contient ne doivent pas être redondantes avec le texte. Cependant, pour les données concernant les résultats principaux de l'étude, il est normal de les énoncer de façon synthétique dans le texte et de les faire figurer également avec tous les détails nécessaires dans le tableau.

2. FIGURES

Elles doivent être claires, compréhensibles et crédibles. Elles servent à représenter des résultats importants. Les critères de qualité d'une figure à vérifier concernent les points suivants :

1. Y a-t-il un titre informatif à la figure ? Une légende explicative ?
2. La figure est-elle bien référencée dans le texte ?
3. Les axes sont-ils sur la figure ? Sont-ils bien référencés ? Les échelles et les divisions sont-elles bien marquées ? Les unités sont-elles précisées ? Les minimums et maximums choisis sur chaque axe sont-ils pertinents ?

4. Le nombre de figures est-il raisonnable ? N'y a-t-il pas de redondances d'information avec le texte?
5. Les figures portant sur les mêmes grandeurs sont-elles toutes à la même échelle pour éviter les distorsions de lecture?
6. Les effectifs sont-ils mentionnés ? La variabilité est-elle représentée (écart-type, intervalle de confiance...). Les valeurs statistiquement significatives sont-elles clairement identifiables ?

OBJECTIF 12°: VERIFIER LA PRESENCE DES INDICES DE DISPERSION PERMETTANT D'EVALUER LA VARIABILITE DES MESURES ET LEURS ESTIMATEURS

1. LES QUESTIONS A SE POSER

- Tous les résultats concernant les critères de jugement sont-ils exprimés avec les intervalles de confiance voire le paramètre de dispersion approprié ?
- Pour la description initiale de l'échantillon concernant les variables quantitatives, a-t-on choisi correctement la présentation de la valeur centrale de la distribution (moyenne ou médiane) et de sa dispersion (écart type, quartiles..) ?
- Le choix de la représentation des résultats par des figures ou des tableaux est-il judicieux ?

2. OU CHERCHER LA PRESENCE DES INDICES DE DISPERSION ?

Dans la section résultats sans oublier les tableaux et figures.

3. RAPPELS THEORIQUES : INDICES DE DISPERSION ET INTERVALLE DE CONFIANCE

Un indice de dispersion renseigne sur l'étalement de la distribution d'une variable. Il complète le paramètre de position et dépend de la nature de la variable.

Quand on estime un critère de jugement, on doit renseigner la précision de l'estimation par un indice de dispersion ou par l'intervalle de confiance à 95%. Pour chaque type de variable, l'estimateur et l'indice renseignant la précision sont reportés dans le tableau ci-dessous.

Variables, estimateurs et indices renseignant la précision

Variable	Estimateur	Indice de dispersion /intervalle de confiance	Exemples
Qualitative	pourcentage	IC 95%	
Quantitative	Moyenne (distribution gaussienne)	Ecart-type (= SD) (distribution gaussienne) IC 95%	Age: moyenne, (écart-type) : 61,3 (11,3) (N=299)
	Médiane (distribution non gaussienne)	Interquartile, (Q1-Q3)=P25 - P75 min - max	Age: médiane (Q1 - Q3) : 25,7 [15 -47]
Prévalence incidence	Estimation ponctuelle	IC 95%	Prévalence de l'asthme induit par l'effort: 9,9% IC95%: [8,2% – 11,7%]
Sensibilité Spécificité VPP, VPN	Estimation ponctuelle	IC 95%	Sensibilité du scanner multicoupe vs coronarographie, Nb/Tot (% [IC95%]): 149/157 (IC95%, [8,2% – 11,7%])
OR RR, RRI HR	Estimation ponctuelle	IC 95%	RRI de Tachyarythmie auriculaire de l'amiodarone versus placebo RRI= 0,52 IC95% [0,34 - 0,69]

Abréviations : IC 95% : intervalle de confiance à 95%; VPP : valeur prédictive positive; VPN : valeur prédictive négative; OR : odds ratio ou rapport de cotes ; RR : risque relatif; RRI ou HR : rapport de risques instantanés ou hazard ratio; SD : déviation standard ; Min : minimum; Max : maximum; Interquartile : les valeurs du premier quartile (Q1) ou du 25ème percentile (P25) et du troisième quartile (Q3) ou du 75ème percentile (P75) sont données

Remarque : Les indices reportés sur même ligne doivent être associés.

Exemple : un interquartile ne doit pas accompagner une moyenne.

Ecart-type de la moyenne (SEM) - Attention: ne pas confondre l'écart-type de la moyenne (SEM) et

l'écart-type (SD). $SEM = SD / \sqrt{N}$

où N est l'effectif et SD la déviation standard ou écart-type.

Si N=100, SEM est 10 fois plus petit que l'écart-type. La dispersion de la variable est pourtant toujours la même car elle est donnée par l'écart-type.

OBJECTIF 13°: DISCUTER LA NATURE ET LA PRECISION DES CRITERES DE JUGEMENT DES RESULTATS

1. LES QUESTIONS A SE POSER : CRITERE DE JUGEMENT

A. Sa définition est-elle précise ?

Le moment où le critère de jugement (CJ) est recueilli et sa méthode d'évaluation sont-ils précisés ?

B. Est-il pertinent pour répondre à l'objectif de l'étude ?

- Est-il clinique, intermédiaire ou de substitution ?
- Le moment de son recueil est-il pertinent ?
- Est-il unique simple, unique composite ou multiple ?
- Le critère de jugement principal (CJP) est-il unique et pertinent ? (à défaut : consensuel ?)
- S'il est unique composite, ses composantes ont-elles toutes la même pertinence clinique ?
- Si le CJP est multiple, les critères ont-ils tous la même pertinence clinique ?
- S'il y a plusieurs CJ, sont-ils hiérarchisés ?

C. Est-il sujet à des biais de classement/information ?

- Sa mesure est-elle standardisée ? (identique pour tous les sujets)
- Est-il objectif ou subjectif ?
- S'il est subjectif, est-il validé ? (référence dans le texte à une étude de validité)
- Est-il fiable et reproductible ?
- Le recueil du critère de jugement est-il fait en insu du groupe d'étude ?

D. Sa précision est-elle adéquate ?

- Le nombre de sujets nécessaire a-t-il été calculé pour le CJP ?
- L'intervalle de confiance (IC) au risque alpha (alpha est défini pour le calcul du NSN dans la section méthodes, généralement alpha = 5% et IC à 95%) est-il donné pour le CJP ?
- Si les critères de jugement ne sont pas hiérarchisés ou si le CJP est multiple, a-t-on précisé une règle de décision pour conclure au niveau statistique ?

2. OU CHERCHER LES INFORMATIONS: RELATIVES A LA NATURE ET A LA PRECISION DES CRITERES DE JUGEMENT DES RESULTATS ?

- Pour la nature des critères de jugement, dans la section méthodes et parfois dans la section introduction (objectif)
- Pour la précision des critères de jugement, dans les sections méthodes et résultats

3. RAPPELS THEORIQUES :

3.1. DEFINITION (GLOSSAIRE CNCI)

Critère de jugement : variable observée et/ou mesurée dont l'interprétation va permettre de répondre à la question posée dans l'objectif.

Dans un essai clinique, il s'agit du critère qui permet d'évaluer l'effet du traitement ou d'une intervention.

Dans une enquête épidémiologique, il s'agit d'un événement ou de la survenue d'un événement. (enquête exposés/non exposés), ou d'une exposition (enquête cas-témoins), d'un facteur pronostique.

Toujours discuter la nature d'un critère de jugement en fonction des objectifs de l'étude.

3.2. CARACTERISTIQUES

Nature

❶	Unique simple	<i>survenue de décès dans les 6 mois post randomisation</i>
	Unique composite	<i>survenue de décès ou d'infarctus du myocarde dans les 6 mois post randomisation</i>
	Multiple	<i>score d'incapacité à l'échelle AVQ-B et score à l'échelle AVQ-D</i>
❷	Objectif	<i>décès, poids</i>
	Subjectif	<i>échelle de qualité de vie SF36</i>
❸	Clinique	critère de jugement correspondant à des objectifs thérapeutiques (objectif de guérison (maladie bénigne), de prévention primaire ou secondaire, symptomatologique, amélioration de la qualité de vie). <i>Ex: survenue d'infarctus du myocarde, décès</i>
	Intermédiaire	critère de jugement documentant les mécanismes d'action du traitement <i>Ex: paramètres biologiques ou physiologiques, pression artérielle</i> <i>Ex taille des cancers dépistés dans l'évaluation d'un dépistage</i>
	De substitution	examen de laboratoire ou signe physique utilisé à la place d'un critère clinique. Il doit être très bien « corrélé » au critère clinique ou prédictif du critère clinique. <i>Ex: volume d'expiration forcée au lieu de brachygnée</i>

Exemples

<i>Pathologie</i>	<i>Critère intermédiaire</i>	<i>Critère clinique</i>
<i>Hypertension artérielle</i>	<i>Pression artérielle</i>	<i>Evènements coronariens, AVC</i>
<i>Dépression</i>	<i>Score de dépression</i>	<i>Normalisation du score de dépression</i>
<i>Rhumatisme inflammatoire</i>	<i>Vitesse de sédimentation</i>	<i>Handicap fonctionnel, déformation</i>
<i>Ostéoporose</i>	<i>Densité osseuse</i>	<i>Fracture du col fémoral</i>
<i>Infarctus du myocarde</i>	<i>Taux de repermeabilisation coronaire</i>	<i>Décès, évènements coronariens</i>
<i>Diabète non insulino-dépendant</i>	<i>Glycémie, hémoglobine glycosylée</i>	<i>Evènements cardiovasculaires, rétinopathie, insuffisance rénale ..</i>

Hiérarchisation des critères de jugement

On distingue le critère de jugement principal ou primaire (CJP) des critères de jugement secondaires (CJS).

L'étude vise à répondre à l'objectif principal exprimé avec le critère de jugement principal.

Le calcul du nombre de sujets nécessaire est basé sur le critère de jugement principal.

Pertinence d'un critère de jugement

Un critère est pertinent s'il permet de répondre à l'objectif de l'étude.

Ce tableau regroupe les critères de jugement attendus et les paramètres statistiques associés en fonction du type de question

Question	Critère de jugement & paramètre statistique associé
Epidémiologie descriptive	Pathologie : Incidence d'une pathologie
	Pathologie : Prévalence d'une pathologie
Evaluation de traitement	Efficacité: survenue d'un événement mesurant l'efficacité; incidence de l'événement, qualité de vie : score sur échelle spécifique
	Sécurité: effets secondaires; incidence des effets secondaires.
Evaluation de procédure diagnostique	Fiabilité: Signe, pathologie; indice de concordance (ex: kappa)
	Validité: Signe, pathologie; sensibilité, spécificité, rapport de vraisemblance, valeurs prédictives, valeur seuil pour les tests quantitatifs / gold standard
Evaluation d'une procédure de dépistage	Qualité : sensibilité, spécificité, valeurs prédictives de la <i>procédure</i> (test+bilan diagnostique) Efficacité : critère indépendant de l'avance au diagnostic (ex : mortalité et non survie dans le dépistage des cancers)
Epidémiologie analytique, Recherche de facteur de risque	Survenue de maladie; Incidence maladie chez exposés et non exposés (cohorte)
Estimation d'un pronostic	Survenue (incidence) de décès, complication, rechute; (Variable dépendant du temps)

Si on recherche une pertinence clinique (recherche clinique à visée pragmatique, épidémiologie) le CJP doit être clinique ou correspondre à des marqueurs biologiques de l'évolution de la maladie (*ex taux de CD4 dans le VIH, ...*)

En recherche clinique à visée explicative, le CJP peut être intermédiaire. Il faut alors s'assurer de la relation de transitivité entre ce critère intermédiaire et un critère clinique.

Si le CJP est unique composite, il faut que ses composantes aient la même pertinence clinique
Contre exemple: décès, infarctus du myocarde, hospitalisation

Critères de jugement et biais

L'évaluation d'un critère de jugement est sujet à des biais de classement (cf objectif 14 - biais).

Un CJ objectif est fiable et reproductible. Il est très peu sujet à des biais de classement. Un CJ subjectif est sujet à des biais de classement.

Pour limiter le biais de classement il est nécessaire que le CJ soit validé (c'est à dire une étude a évalué sa bonne fiabilité ou sa bonne reproductibilité), le biais de classement est alors limité.

Le biais de classement est différentiel si l'évaluation n'est pas faite en insu du groupe étudié.

Le biais de classement est non différentiel si l'évaluation est faite en insu du groupe étudié.

Critères de jugement et précision

A. Précision de la définition d'un critère de jugement

- La définition d'un critère de jugement doit être précise : elle contient une variable, sa méthode de mesure et le moment où elle est évaluée
- *Exemple : acquisition du VIH (évaluée par PCR ADN) entre les âges de 7 jours et 38 semaines*
- *Si le CJP est la résultante d'un calcul particulier, celui-ci doit être fourni en annexe.*

B. Précision d'un critère de jugement

- Le nombre de sujets nécessaire est calculé pour le critère de jugement principal. Vérifier que le nombre de sujets recrutés (ou d'évènements) correspond aux hypothèses de calcul du NSN. Remarque : Un critère de jugement fréquent permet de réduire le NSN.
- Si l'analyse du CJ est à visée de confirmation, on doit préciser la règle de décision relative au risque alpha pour conclure au niveau statistique (Le risque alpha est précisé). On ne corrige pas alpha si l'analyse est à visée exploratoire. Généralement, l'analyse principale portant sur le CJP est à visée de confirmation.
- Le résultat doit être exprimé avec l'estimation ponctuelle accompagnée de l'intervalle de confiance au seuil alpha (généralement à IC 95%)
- Dans le cas d'un CJP multiple, la règle de décision relative au risque alpha pour conclure au niveau statistique doit être précisée. Exemple de règle : dans le cadre d'un essai thérapeutique avec plusieurs CJP, on ne peut conclure à l'efficacité du traitement que si tous les petits $p < 0,05$. Ainsi, si un seul petit p est $> 0,05$, on ne peut pas conclure à l'efficacité du traitement.
- Les analyses des CJS sont exploratoires sauf mention contraire.

La conclusion ne porte que sur le CJP. Les analyses des CJS sont exploratoires sauf mention contraire. Elles viennent en complément de l'analyse principale du CJP.

Le CJP doit être le plus pertinent possible. Dans un essai clinique, c'est l'intérêt du patient qui prime et non l'intérêt du traitement.

OBJECTIF 14°: RELEVER LES BIAIS QUI ONT ETE DISCUTES, RECHERCHER D'AUTRES BIAIS D'INFORMATION ET DE SELECTION EVENTUELS NON PRIS EN COMPTE DANS LA DISCUSSION ET RELEVER LEURS CONSEQUENCES DANS L'ANALYSE DES RESULTATS

1. LES QUESTIONS A SE POSER

- a. Les auteurs soulignent-ils dans la discussion un ou des biais à prendre en compte dans l'interprétation des résultats ? ont-ils cherché à en déterminer le sens (sur/sous-estimation du paramètre) ?
- b. Ont-ils essayé de les contrôler ou d'en atténuer l'effet par le protocole mis en place et/ou les ajustements à l'analyse?
- c. Existe t-il d'autres biais omis par les auteurs ?
- d. L'ensemble des biais identifiés (atténués ou non, explicités ou non) invalide t-il les résultats de l'étude? Donner leur nature, le sens du biais et quantifier le biais si possible

2. OU CHERCHER LES INFORMATIONS RELATIVES AUX BIAIS ?

- e. Pour relever les biais qui ont été discutés, dans la section discussion (souvent paragraphe limites de l'étude)
- f. Pour rechercher d'autres biais éventuels non pris en compte dans la discussion et relever leurs conséquences dans l'analyse des résultats, dans les sections méthodes et résultats

3. RAPPELS THEORIQUES

3.1. GENERALITES SUR LES BIAIS

Définition (glossaire CNCI)

Un biais est une erreur systématique qui fausse les résultats dans un sens donné .

Attention : ne pas confondre un biais avec une erreur aléatoire (fluctuation d'échantillonnage qui débouche sur une imprécision) ou avec une erreur dans une conclusion statistique (risque alpha et risque bêta ou manque de puissance).

Pour tout biais, donner

- le nom de sa famille,
- son sens (surestimation ou sous-estimation, si possible) et
- le quantifier (biais négligeable, faible, modéré, important, si possible).

On classe les biais en 3 familles (glossaire CNCI) selon leur cause

- Biais de sélection
- Biais de classement
- Biais de confusion

Deux types d'erreurs de classement: biais différentiel et biais non différentiel

- **Un biais est différentiel** si les erreurs sur les informations recueillies affectent différemment les groupes soumis à comparaison. Le sens du biais est imprévisible a priori : on a soit une surestimation soit une sous-estimation de l'association.

Exemple : l'évaluation de l'exposition des sujets dans une enquête cas-témoin (non en aveugle) faite en connaissant le statut de la maladie

- **Un biais est non différentiel** si les erreurs sur les informations recueillies affectent indifféremment les groupes soumis à comparaison. Ces erreurs sont statistiquement indépendantes du groupe de comparaison. Le sens du biais est prévisible : on a toujours une diminution de l'association entre le facteur étudié et le critère de jugement : OR tend vers 1; RR vers 1; HR vers 1

Exemple : évaluation de la consommation d'alcool en insu du traitement reçu dans un essai randomisé - évaluation erronée de l'exposition à un facteur pour des questions liées à l'outil de mesure (se produit quels que soient les sujets)

3.2. FAMILLE DES BIAIS DE SELECTION

Définition (glossaire CNCI)

Biais dans la constitution de l'échantillon, qui va se retrouver non représentatif de la population cible pour des facteurs liés au problème étudié.

Attention : il s'agit de l'échantillon analysé.

Processus

Un biais de sélection peut survenir à 2 niveaux :

- **Au niveau de la constitution de l'échantillon.** Il s'agit d'un échantillon non représentatif de la population source ou de la population à laquelle on veut extrapoler les résultats
Exemple : sujets âgés de 70 ans ou plus sans co-morbidité ne sont pas représentatifs de la population des sujets âgés
- **Au niveau des groupes analysés :** les groupes analysés ne diffèrent pas seulement par le facteur étudié mais aussi par un autre facteur (facteur pronostique) qui peut modifier les résultats.
Exemple : essai contrôlé randomisé avec 10% de sorties d'étude liées à des effets indésirables dans un groupe de traitement => résultats d'efficacité biaisés

Biais inclus dans cette famille (noms des biais non au programme des ECN)

- **Biais d'attribution** (allocation bias) : biais survenant quand les sujets ne sont pas répartis aléatoirement dans les groupes étudiés
Exemple : procédure de randomisation détournée
- **Biais d'attrition** : biais survenant quand on exclut de la population d'analyse des sujets initialement inclus dans l'étude et que ces exclusions sont liées au résultat
Exemple : dans un essai randomisé, analyse non en intention de traiter et motifs des exclusions liées au résultat
- **Biais de perdus de vue** : biais survenant quand les motifs des pertes de vue sont liés au résultat.

Comment contrôler les biais de sélection

On peut les contrôler à deux niveaux : lors de la planification de l'étude ou lors de l'analyse

- **Planification**

Recrutement par tirage au sort	concerne surtout les enquêtes d'observation <i>Exemples : enquête descriptive : sondage aléatoire; enquête cas-témoin : sélection par tirage au sort des témoins dans population d'où sont issus les cas</i>
Recrutement exhaustif	<i>Exemple : enquête cas-témoin : recrutement systématique de tous les cas définis par des critères diagnostiques valides et appliqués à tous les sujets</i> <i>Exemple: enquête de cohorte, essai clinique, évaluation procédure diagnostique : recrutement systématique de tous les patients éligibles sur une période donnée</i>
Randomisation du facteur étudié	si l'objectif de l'étude le permet. S'assurer que la randomisation est bien faite (cf objectif 5)
Suivi complet et/ou relances	concerne toutes les études. Pour éviter les perdus de vue dans études prospectives et les données manquantes dans toutes les études

- **Analyse (limite le biais mais ne parvient pas à le neutraliser totalement)**

Comparaison initiale des groupes	Les facteurs pronostiques doivent être similaires (comparabilité clinique et non statistique). Sinon, ajustement sur les facteurs pronostiques déséquilibrés entre les groupes.
Analyse en intention de traiter (ITT) /analyse de tous les sujets inclus dans leur groupe initial de randomisation	L'analyse en ITT concerne uniquement les essais cliniques. Elle n'élimine pas le biais si le suivi des patients est incomplet et si les motifs de sortie d'étude ou de perdus de vue sont liés aux résultats
Analyse en intention de dépister	Même commentaire que pour l'analyse en ITT
Ajustement sur facteurs	Il s'agit des facteurs pronostiques déséquilibrés, des facteurs sur lesquels la randomisation a été stratifiée (centre ...) et les facteurs d'appariement (enquêtes cas-témoins, enquêtes de cohorte) Remarque : si les groupes analysés sont très différents, un ajustement ne suffit pas à neutraliser le biais car on ne peut pas ajuster sur tous les facteurs pronostiques (tous ne sont pas recueillis)
Analyse de sensibilité	Elle permet de déterminer le sens du biais et de le quantifier dans la mesure du possible <i>Exemple : analyse de la robustesse des résultats si les perdus de vue sont liés au facteur étudié ou à la réponse.</i>

3.3. LA FAMILLE DES BIAIS DE CLASSEMENT

Définition (glossaire CNCI)

Biais dans la mesure du facteur de risque ou dans la certitude de la maladie. Cette erreur est quasi inévitable puisque aucun outil de mesure (interrogatoire, examen, test) n'est parfait.

Ce biais peut toucher aussi bien le facteur étudié (par exemple l'exposition dans une enquête de cohorte exposés-non exposés) que les critères de jugement (exemple : score d'un questionnaire, décès pour cause spécifique).

Processus

Un biais de classement peut résulter de

- la subjectivité de l'enquêteur : variabilité
- la subjectivité des enquêtés : mémorisation, refus de répondre, déni
- la méthode de mesure : rythme de suivi différent, temps de mesure différent, outil non valide

Biais inclus dans cette famille (noms des biais non au programme des ECN)

- **Biais d'information/mesure** : biais survenant quand le protocole de mesure n'est pas standardisé (le même pour tous) et/ou est sujet à interprétation.
Exemple : questionnaire de qualité de vie non validé
- **Biais de mémorisation (recall bias)** : biais survenant lors du rappel de l'exposition ou du critère de jugement dans la vie passée.
Exemple : enquête cas-témoin : les cas peuvent faire plus d'effort pour se souvenir d'une exposition passée que les témoins
- **Biais d'évaluation** : biais survenant quand des facteurs subjectifs externes influencent l'évaluation du critère de jugement ou de l'exposition.
Exemple : dans toute étude, ce biais survient quand l'évaluation du critère de jugement ou de l'exposition n'est pas faite en aveugle. Il est moindre quand le critère est objectif.

Comment contrôler les biais de classement

Les moyens à disposition pour contrôler les biais de classement doivent être prévus lors de la planification de l'étude

Les règles suivantes s'appliquent au facteur étudié et au critère de jugement et à toutes les études

- **validité et fiabilité des mesures :**
 - ✓ la mesure doit être précise, exacte et reproductible (reproductibilité intra-observateur et inter-observateur).
 - ✓ Pour les critères subjectifs, il faut une référence dans le texte à une étude de validation; à défaut, des résultats de reproductibilité
Exemple : questionnaire de qualité de vie (subjectif) doit avoir été validé. Vérifier qu'il y a une référence dans le texte à une étude de validation.
- **Standardisation des procédures**
 - ✓ Les mêmes procédures doivent être appliquées à tous les sujets de l'étude: même procédé de recueil de mesure, même rythme de mesure (si prospectif), même définition du facteur étudié et des critères de jugement pour tous les sujets.
Exemples : mêmes critères diagnostiques appliqués à tous les sujets dans une enquête cas-témoin, à chaque fois que c'est possible même contenu et rythme de suivi dans les enquêtes exposés/non exposés
- **Evaluation/recueil des données en aveugle/en insu.**
 - ✓ Essais cliniques : évaluation des critères de jugement en insu du traitement
 - ✓ Résultat d'un examen diagnostique en insu du résultat du gold standard et vice versa
 - ✓ Enquête cas-témoin : évaluation de l'exposition en insu du statut malade/non malade du sujet
 - ✓ Enquête de cohorte exposés non exposés : évaluation de la maladie en insu de l'exposition
- **Formation des enquêteurs**
Pour avoir une standardisation des procédures

3.4. LA FAMILLE DES BIAIS DE CONFUSION

Définition (glossaire CNCI)

Biais provoqué par un facteur de confusion interagissant avec le facteur de risque étudié dans l'étude du lien entre ce facteur et la maladie.

Un facteur F joue le rôle de facteur de confusion dans la relation entre E et M si dans la population cible

- la relation brute entre E et M n'est pas la même que celle obtenue en tenant compte de F (relation ajustée).
- et F est lié à M et F est lié à E mais F n'est pas une conséquence de E

Un facteur de confusion existe au niveau de la population avant même de constituer l'échantillon de l'étude.

Comment neutraliser les facteurs de confusion

On peut neutraliser les facteurs de confusion lors de la planification de l'étude et/ou lors de l'analyse, mais cette neutralisation n'est jamais complète.

- **Planification**

tirage au sort du facteur étudié (randomisation)	=> comparabilité des groupes <i>Exemple : essais thérapeutiques.</i> Cette règle n'est pas applicable à toutes les études.
restriction de la population d'étude à certaines catégories	<i>Exemple : exclusion de patients atteints de bilharziose (facteur de risque connu de cancer de la vessie) dans une enquête étiologique sur le cancer de la vessie en Europe. Les cas de bilharziose sont rares en Europe. Leur exclusion ne met pas en péril l'extrapolation des résultats.</i>
appariement sur les facteurs de confusion voire de risque de la maladie déjà connus	<i>En général âge, sexe ...</i> enquêtes étiologiques Attention: ne pas confondre facteur de risque et facteur de confusion un sur-appariement peut masquer une liaison

- **Analyse**

ajustement sur les facteurs de confusion	<p>On utilise des modèles multivariés (voir cours de statistiques) qui donnent l'effet propre du facteur étudié indépendamment des autres co-variables entrées dans le modèle. On obtient un OR, RR ou HR ajusté sur les facteurs de confusion entrés dans le modèle multivarié.</p> <p>Attention: ceci nécessite d'avoir prévu dans le protocole de recueillir les facteurs de confusion.</p>
--	--

Remarque :

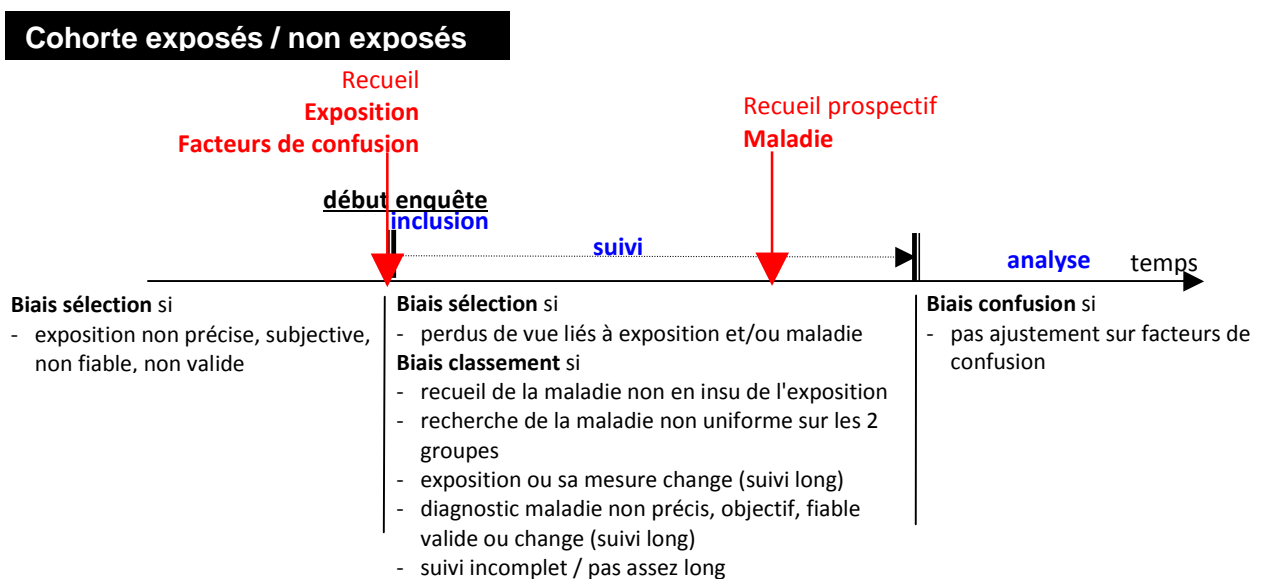
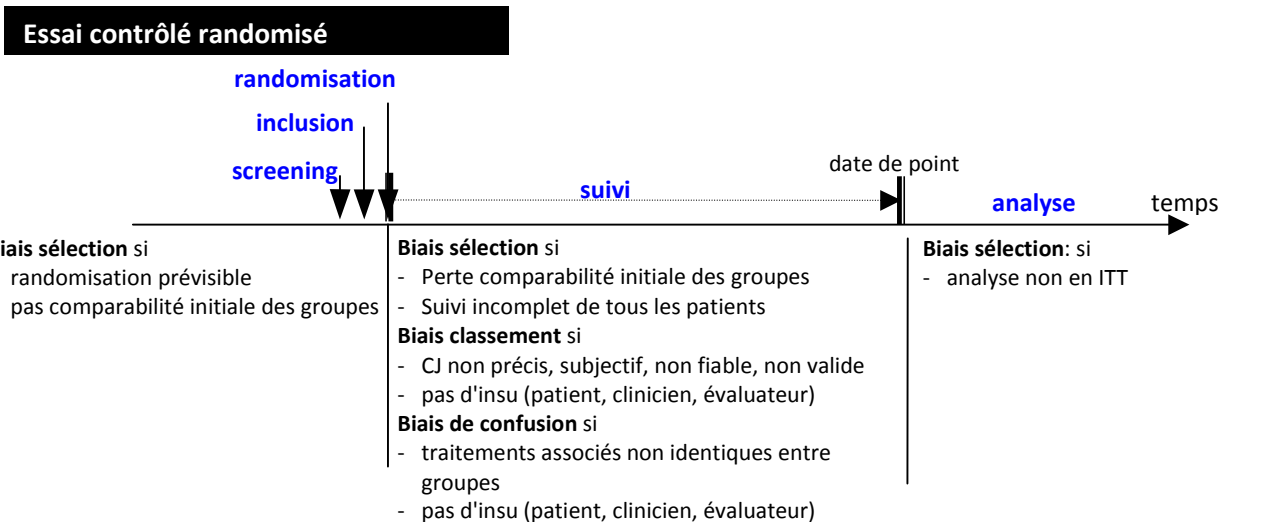
Certaines situations relèvent de plusieurs familles de biais

Exemples : dans une enquête cas-témoins où le recrutement se déroule sur une assez longue période avec une évolution des critères diagnostiques, on a à la fois un biais de classement (un sujet classé au départ non malade aurait été classé ensuite malade) et un biais de sélection car on est à la phase de constitution de l'échantillon

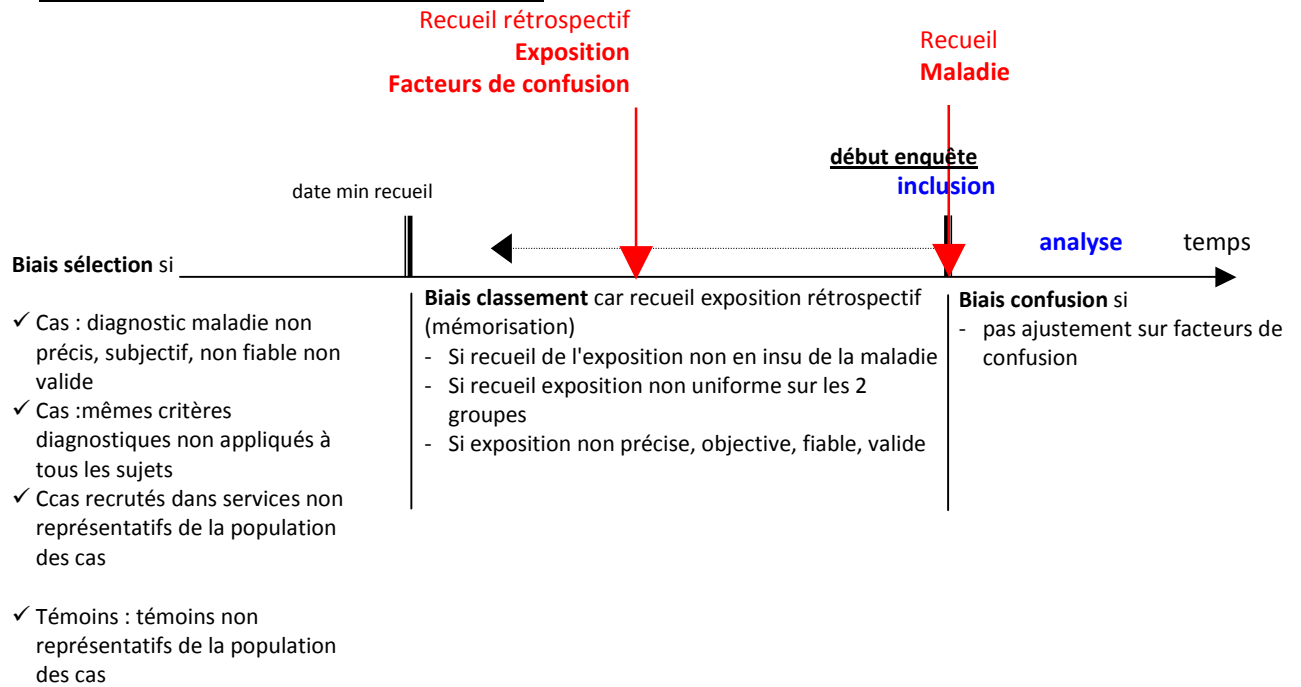
Les moyens même mis en place pour prévenir les biais ne garantissent pas l'absence de biais et ne dispensent pas les auteurs d'en discuter l'existence.

3.5. PLACE DES BIAIS DANS LES ETUDES

Les biais peuvent se situer à chaque stade de l'étude : lors de la planification, lors de la réalisation et lors de l'analyse des données



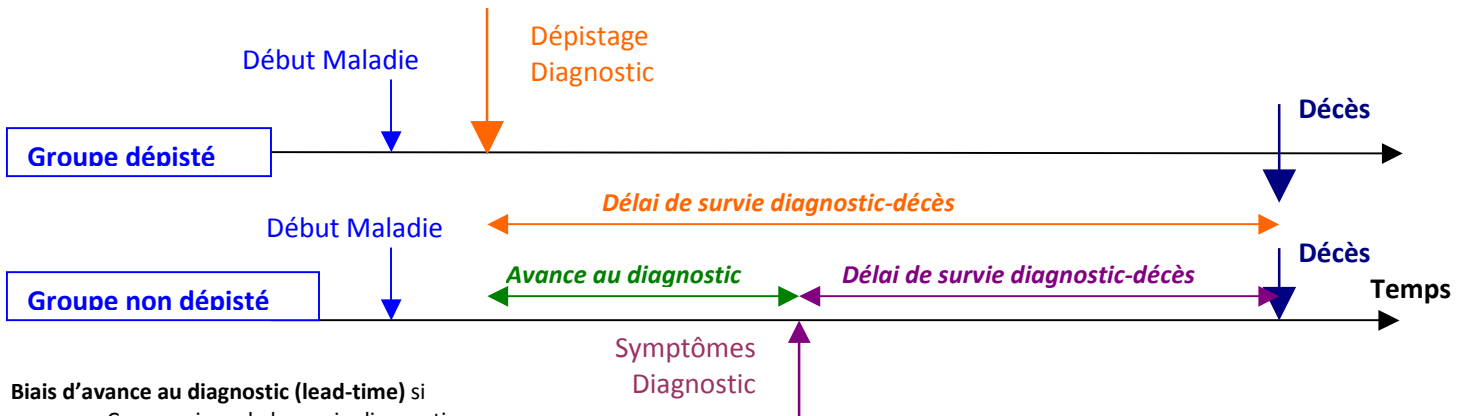
Cas - Témoins



Evaluation de procédure diagnostique - Valeur diagnostique Grille de lecture - Validité des résultats – 8 points clés

	Oui	Non
• Étude prospective voire transversale ? Les deux examens évaluent le même état du sujet		Biais de classement
• Échantillon de patients représentatif de la population cible (prévalence et sévérité de la pathologie) ?		Biais de sélection
• Définition parfaite de la maladie ?		Biais de sélection
• Test référence (gold standard) valide et reconnu ?		Biais de classement
• Test de référence déterminé <i>a priori</i> (indépendant du test étudié) ?		Biais de classement
• Test étudié fiable et ses conditions d'utilisation décrites avec précision (positivité définie) ?		Biais de classement
• Insu des deux tests ?		Biais de classement
• Tous les sujets subissent les deux tests ?		Biais de sélection

Essai randomisé prospectif appliqué au dépistage : biais spécifiques

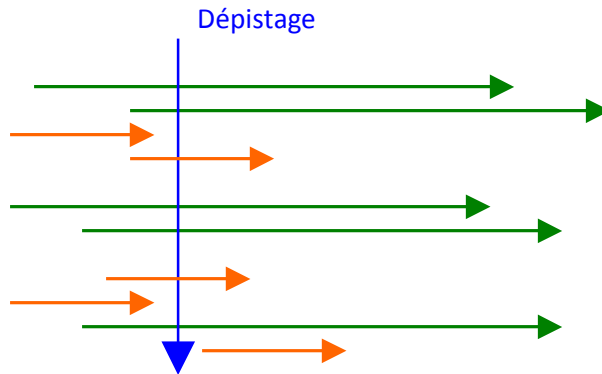


Biais d'avance au diagnostic (lead-time) si

- Comparaison de la survie diagnostic-décès : toujours plus élevée dans le groupe dépistage même s'il est inefficace (date du décès inchangée)

Biais lié à la vitesse d'évolutivité (length-time)

- tumeurs lentes $\frac{5}{5}$ des sont identifiées par le dépistage
- tumeurs rapides $\frac{2}{5}$ des sont identifiées par le dépistage
- ⇒
- Taux plus élevé dans le groupe dépisté de formes à évolution lente de bon pronostic : effet sur le taux de mortalité



Autres biais survenant dans les évaluations de dépistages

- biais de « sur diagnostic »
- biais de sélection liés aux sujets par définition volontaires
- biais d'indication : certains sujets sont adressés au dépistage parce qu'ils ont des symptômes

OBJECTIF 15 : VERIFIER LA LOGIQUE DE LA DISCUSSION ET SA STRUCTURE, RECONNAITRE CE QUI RELEVE DES DONNEES DE LA LITTERATURE ET CE QUI EST L'OPINION PERSONNELLE DE L'AUTEUR

1. LES QUESTIONS A SE POSER :

- a. Les auteurs apportent-ils une réponse à la question posée ?
- b. L'interprétation des résultats observés dans l'étude est-elle rationnelle ?
- c. Les points forts de l'étude sont-ils exposés ?
- d. Les points faibles et les limites de l'étude sont-ils répertoriés ? Sont-ils appropriés ?
- e. Les auteurs ont-ils confronté leurs résultats à la littérature récente ?
- f. Ce qui relève de la littérature a-t-il été référencé ?
- g. Le caractère généralisable des résultats de l'étude a-t-il été discuté ? Si les auteurs concluent qu'on peut extrapoler les résultats à telle population, est-ce justifié ?
- h. La conclusion des auteurs est-elle étayée par les résultats de l'étude ?
- i. Y a-t-il des pistes de recherche complémentaire proposées ?
- j. N'y a-t-il pas des erreurs de forme ?
 - ✓ Discussion trop longue, revue de la littérature injustifiée (non en rapport direct avec les résultats de l'étude)
 - ✓ présentation de nouveaux résultats non présentés dans le chapitre résultats (par exemple résultats d'analyse en sous-groupes)...

2. OU CHERCHER LES INFORMATIONS RELATIVES A LA DISCUSSION ?

- dans la section discussion (la section des résultats ne doit pas faire l'objet d'interprétation/ discussion)

3. RAPPELS THEORIQUES

Une discussion doit présenter les 6 points suivants. Généralement elle suit la logique suivante mais ce n'est pas une règle (cf objectif 22).

1. Résumé des principaux résultats
2. Discussion de leur validité : points forts et points faibles
3. Discussion de leur caractère généralisable
4. Mise en perspective de cette réponse : confrontation des résultats à la littérature
5. Conclusion sur la réponse à la question posée : conclusion clinique
6. Implications et ouverture, perspectives

OBJECTIF 16°: DISCUTER LA SIGNIFICATION STATISTIQUE DES RESULTATS.

1. LES QUESTIONS A SE POSER

- Les tests statistiques sont-ils adaptés à la question posée ?
- L'interprétation de leurs résultats est-elle adéquate ?
- La valeur du « p » est-elle interprétée correctement ?
- Y a-t-il une signification clinique associée à la signification statistique ?
- Le nombre de tests réalisés est-il justifié ?
- En cas d'analyses en sous-groupes, sont-elles licites et leur interprétation est-elle adéquate? (cf objectif 9)

2. OU CHERCHER LES INFORMATIONS

- dans la partie méthodes, paragraphe analyse statistique pour vérifier leur adéquation à la question posée et la justification des études de sous-groupes
- dans la partie résultats, souvent dans les tableaux pour juger de la signification statistique des résultats.
- Parfois en annexe (indices ou tests particuliers)

3. RAPPELS THEORIQUES

A. Risque alpha et risque beta

La signification statistique exprime une probabilité. Elle varie en fonction du type de test utilisé et est dépendante de l'effectif sur lequel est réalisé le test. Ainsi toute différence, aussi petite soit-elle, peut devenir statistiquement significative, si l'effectif sur lequel elle est calculée est suffisamment grand.

Le risque alpha ou risque de 1ère espèce exprime le risque de déclarer une différence significative (sous l'effet d'un traitement ou d'autres facteurs extérieurs) alors qu'elle n'est que le résultat de fluctuations aléatoires. Ce risque est jugé acceptable s'il est inférieur à 5%. (ce qui revient à un taux maximal de faux positifs toléré : 5%). Autrement dit, c'est le risque de rejeter à tort l'hypothèse nulle et de conclure par exemple qu'un traitement est efficace alors qu'il ne l'est pas.

Le risque bêta ou risque de 2ème espèce exprime le risque de déclarer une différence non significative alors qu'elle traduit en fait l'effet d'un traitement ou d'autres facteurs extérieurs. Il est encore appelé manque de puissance et est jugé acceptable s'il est inférieur à 20% (ce qui revient à un taux maximal de faux négatifs toléré : 20%), voire plus souvent 10% dans les essais thérapeutiques de phase III). Autrement dit, c'est le risque de ne pas rejeter l'hypothèse nulle alors qu'elle est vraie et de conclure par exemple qu'un traitement n'est pas efficace alors qu'il l'est.

Un même ensemble de données peut être analysé de plusieurs manières, dont certaines peuvent ne pas être strictement appropriées. Les points suivants sont donc essentiels :

- Les auteurs doivent indiquer quelle méthode statistique a été employée pour chaque analyse (présence d'un chapitre « considérations statistiques » dans la partie « méthodes »)
- Quand on décrit les caractéristiques de chaque groupe soumis à comparaison, on recherche une comparabilité clinique et non statistique car les effectifs ont été calculés pour détecter une différence sur le critère de jugement principal et non sur les caractéristiques initiales. Faire des tests statistiques n'est pas recommandé dans ce cas. L'absence de significativité du test ne doit pas conduire à négliger une différence si elle est jugée importante sur le plan clinique
- Une valeur de p très petite ne signifie pas forcément qu'il existe une grande différence entre les groupes de comparaison : ne pas confondre degré de signification statistique (p) et quantité d'effet (ampleur de la différence entre les groupes : modification de la moyenne ou du taux d'évènements). Une valeur de « p » > 5% ne signifie pas qu'il n'existe pas de différence, mais seulement qu'on n'a pas mis en évidence de différence significative (= non due au hasard). Plusieurs raisons peuvent être à l'origine de cette non signification : l'absence réelle de différence, l'existence d'une différence plus petite que celle qu'on désirait détecter *a priori*, le manque de puissance, un biais...
- Toute différence, aussi petite soit-elle, peut devenir statistiquement significative, si l'effectif sur lequel elle est calculée est suffisamment grand mais cette différence statistiquement significative peut ne pas

être cliniquement significative, c'est-à-dire n'avoir aucun intérêt pour la santé des patients (cf objectif 17).

- Plus on multiplie les tests statistiques, plus on a de chances d'avoir un test significatif uniquement par hasard. Donc vérifier que les auteurs n'ont pas réalisé des tests tous azimuts mais sur les critères de jugement annoncés et qu'ils ont utilisé des corrections *ad hoc pour garantir un risque α global de 5%*.
- Il est toujours très tentant d'analyser le résultat dans des sous groupes d'intérêt (hommes/femmes, formes graves/bénignes de la maladie etc.). On court alors trois risques : le manque de puissance par diminution des effectifs, l'observation d'un résultat significatif du seul fait du hasard, en raison de la multiplication des tests et la destruction de la comparabilité initiale des groupes. Les analyses en sous-groupes doivent donc être justifiées et prévues AVANT le début de l'étude de façon à ce que la planification de l'étude en tienne compte (justification des analyses, hypothèse à vérifier, nombre de sujets à inclure, stratification de la randomisation, test de l'interaction, correction du risque alpha). Dans le cas où elles seraient décidées a posteriori ou sur un effectif trop faible, leurs résultats ne doivent être donnés qu'à titre exploratoire et doivent être vérifiés avec d'autres études. Elles doivent de toute façon être en nombre limité.

B. La valeur du « p » est-elle interprétée correctement ?

Les résultats peuvent être évalués en termes de signification statistique. La valeur de « p » représente la probabilité que la différence observée le soit par hasard, quand en réalité il n'y a pas de différence entre les groupes. Ainsi, une valeur de p de 0,01 signifie que l'on a seulement 1 chance sur 100 d'observer au moins cette différence sous le seul effet du hasard, quand il n'y a pas de différence entre les groupes.

Le seuil de signification (risque α) est fixé à 5% par consensus; on conclut, lorsque $p < 5\%$, qu'il y a une différence statistiquement significative : on considère peu probable que la différence soit due au hasard.

Remarque importante : une valeur de p très petite ne signifie pas forcément qu'il existe une grande différence entre les groupes de comparaison. Cela signifie simplement que l'hypothèse nulle (pas de différence) est très peu vraisemblable. Toutes choses égales par ailleurs, la valeur de p sera d'autant plus petite que l'effectif de sujets est grand.

En conclusion :

- $p < 5\%$ = **test significatif**, la différence observée n'est probablement pas due au hasard,
- $p > 5\%$ = **test non significatif**, la différence observée pourrait être due au hasard .:
- **p compris entre 5% (risque α) et 10% : une tendance statistique** est mise en évidence. Pour infirmer ou confirmer une tendance statistique, il aurait fallu recruter plus de sujets et/ou allonger la durée de suivi de l'étude pour avoir plus d'événements.

Mais un résultat statistiquement significatif peut ne pas être cliniquement significatif.

Nota bene : ne pas confondre tendance statistique et test de tendance (trend test).

Pour savoir s'il y a une relation monotone entre un facteur d'exposition analysé en catégories et le risque d'un phénomène (maladie), c'est à dire si la relation se renforce (ou au contraire diminue) avec des catégories du facteur d'exposition plus élevées, on fait un test de tendance.

Exemple : on étudie la relation entre activité physique et incidence de diabète. L'activité physique est classée en 5 classes selon la fréquence. Les rapports de risque instantanés de survenue de diabète (hazard ratio=HR) sont reportés dans la table ci-dessous (The American journal of medicine, 122: 1115-21).

	Rarely/Never	1-3/mo	Once/wk	2-4×/wk	≥5 times/wk	P, Trend
Age-adjusted	1.00	0.79 (0.68-0.93)	0.74 (0.64-0.86)	0.57 (0.50-0.65)	0.43 (0.36-0.52)	<.001

On observe que HR diminue avec les catégories de fréquence d'activité physique plus élevées. La colonne "p, trend" donne le résultat du test de tendance. Il est significatif ($p < 0.05$). On met en évidence une relation inverse entre fréquence de l'activité physique et risque de diabète. Plus l'activité physique est fréquente, plus le risque de diabète diminue.

1. LES QUESTIONS A SE POSER

- La signification clinique est-elle analysée après avoir évalué les biais ?
- La signification clinique est-elle analysée en référence à la « taille d'effet » et non à la signification statistique ?
- La variable choisie pour cette analyse est-elle pertinente : s'agit-il bien de l'une des variables traduisant les objectifs de l'étude (critères de jugement définis) ?
- Le jugement sur la taille de l'effet utilise-t-il l'intervalle de confiance ?
- Le jugement sur la taille de l'effet fait-il référence à des données externes à l'étude (si oui, sont-elles référencées ?)
- Dans les enquêtes observationnelles, peut-on conclure à la causalité ?

Remarque: privilégier le point de vue clinique.

2. OU CHERCHER LES INFORMATIONS

- dans les résultats pour juger de la « taille » de l'effet.
- dans l'introduction, les méthodes (calcul du nombre de sujets nécessaire) et la discussion pour trouver les éléments permettant de juger de l'intérêt de la taille de l'effet.

3. RAPPELS THEORIQUES

A. Signification clinique

La signification clinique exprime un jugement porté sur l'importance au sens de « taille » du résultat obtenu dans l'étude. Elle justifie la décision clinique, diagnostique, pronostique ou thérapeutique. La taille de l'effet fait référence à l'une des quantités résumant l'étude qui quantifie un écart d'effet entre les groupes comparés.

La taille de l'effet est exprimée en bénéfice relatif et/ou bénéfice absolu.

Le bénéfice relatif caractérise le facteur étudié. Pour des résultats binaires (ex : décès oui/non), il est exprimé par le risque relatif, ou le rapport des cotes (odds ratio, OR) selon l'étude. Pour des données censurées (ou de survie), il est exprimé par le rapport de risques instantanés (hazard ratio, HR).

Le bénéfice absolu caractérise un facteur étudié dans un contexte. Pour des résultats binaires ou les données censurées (ex : décès oui/non), il est exprimé par la réduction absolue de risque ($RAR=R_E-R_C$) et/ou par le nombre de sujets à traiter (NNT) pour éviter un accident ($1/RAR$). Pour des variables quantitatives (taux de cholestérol), il est exprimé par la différence des moyennes $[M_E-M_C]$ et/ou par le d de Cohen ou le g de Hodge ($[M_E-M_C]/\text{écart-type combiné}$). Le d de Cohen ou le g de Hodge sont les équivalents du nombre de sujets à traiter pour les variables qualitatives ou censurées.

L'effet du traitement est d'autant plus grand que la RAR est élevée ou que le NNT est faible.

Exemple : un traitement X réduit le risque de décès de 50% par rapport au placebo ($HR=0,5$). C'est le bénéfice relatif.

Si on applique ce traitement X à une population à bas risque (risque de décès 4% dans le groupe placebo) on s'attend à avoir en moyenne un risque de décès de 2%. Le bénéfice absolu est de 2% ($4\%-2\%$). Ceci signifie que si on traite 100 patients à bas risque, on évite en moyenne 2 décès par rapport à un traitement par placebo. Autrement dit, il faut traiter en moyenne 50 patients à bas risque pour éviter un décès.

Si on applique ce traitement X à une population à haut risque (risque de décès 40% dans le groupe placebo) on s'attend à avoir en moyenne un risque de décès de 20%. Le bénéfice absolu est de 20% ($40\%-20\%$). Ceci signifie que si on traite 100 patients à haut risque, on évite en moyenne 20 décès par rapport à un traitement par placebo. Autrement dit, il faut traiter en moyenne 5 patients à haut risque pour éviter un décès.

Pour chaque critère de jugement, les résultats de l'étude doivent au préalable être rapportés sous la forme de la mesure appropriée dans chaque groupe (par exemple, la proportion de participants avec ou sans l'événement, ou la moyenne et l'écart type des mesures).

B. Taille de l'effet

L'estimation de la taille d'effet doit être suffisamment précise (importance de donner son intervalle de confiance) pour pouvoir éliminer le fait que l'effet puisse être petit (borne inférieure de l'intervalle) donc sans intérêt en pratique. La définition du « plus petit effet intéressant » en pratique est arbitraire et difficile. Elle

doit tenir compte de plusieurs paramètres qui sont extérieurs à l'étude et qui doivent être référencés dans la discussion:

- La gravité de la pathologie
 - ✓ Traiter 100 malades pendant 2 ans pour éviter une hospitalisation n'a pas la même « importance » que traiter 100 malades pour éviter 1 décès.
- De la fréquence de la maladie : plus une maladie est fréquente plus un petit bénéfice peut correspondre à un bénéfice substantiel en terme de nombre de sujets concernés par le bénéfice dans la population.
 - ✓ Ainsi, un risque relatif de 1.3 peut être vu comme négligeable au niveau de l'individu, mais important au niveau collectif si ce risque concerne un problème fréquent dans la population.
 - ✓ Une augmentation relative de risque de 30% peut paraître importante. Si elle s'applique à une maladie dont la fréquence est de $1/10^6$, le risque de maladie passe de 1 à $1,3 \times 10^{-6}$, ce qui paraît négligeable...

Ce problème ne concerne pas seulement l'essai contrôlé ou les études de cohorte. On doit aussi se poser cette question pour d'autres études par exemple les études diagnostiques (un test ayant une sensibilité de 90% est-il « un bon test » ?) ou les études analysant la relation entre deux variables (un coefficient de corrélation de 0.90 traduit-il une « forte relation » entre ces 2 variables ?). En résumé, l'adjectif « bon », « mauvais », « fort » ou « faible » qualifiant un résultat doit être justifié.

C. Peut-on conclure à la causalité ?

La causalité ne s'affirme pas du seul fait d'un test « statistiquement » significatif, la causalité peut s'affirmer si, et seulement si :

- il s'agit d'un essai randomisé,
- il n'y a pas de biais dans la randomisation,
- la randomisation n'est pas détruite par un trop grand nombre de perdus de vue, de données manquantes ou de modifications du protocole,
- l'analyse est réalisée en intention de traiter.
- Dans une étude observationnelle, un faisceau d'arguments doit être réuni point D objectif 17.

D. Critères permettant de conclure à la causalité

La causalité ne s'affirme pas du seul fait d'un test « statistiquement » significatif, la causalité peut s'affirmer si, et seulement si :

- ✓ il s'agit d'un essai randomisé,
- ✓ il n'y a pas de biais dans la randomisation ou les groupes sont comparables au départ pour les facteurs pronostiques,
- ✓ la randomisation n'est pas détruite par un trop grand nombre de perdus de vue, de données manquantes ou de modifications du protocole (traitement efficace concomitant),
- ✓ l'analyse est réalisée en intention de traiter.

Dans une enquête observationnelle, la causalité entre le facteur étudié et la maladie est suggérée par un faisceau d'arguments.

Critères de Hill

Critères internes à l'étude

- **Force de l'association** : quand le degré d'association (RR ou OR) est très grand il faudrait un fort effet de confusion pour faire disparaître l'association
- **Cohérence chronologique**: Exposition au facteur doit précéder l'apparition de la maladie : permet d'exclure la relation inverse
- **Relation dose-effet** : si l'exposition augmente, le risque augmente : quand ce critère est présent, il est très spécifique mais il n'est pas vrai dans tous les cas
- **Spécificité de l'association** : **si un facteur de risque est constamment relié uniquement à la maladie étudiée, il apparaît vraisemblable qu'une relation causale existe** : ce critère découle d'une analogie avec les maladies infectieuses mais son absence ne remet pas en question une relation causale

– **Reproductibilité** : Constance de l'association et reproductibilité des résultats de l'étude avec d'autres équipes, en différents lieux, circonstances et temps : c'est un bon critère

• **Critères externes à l'étude**

– **Plausibilité** ou **Cohérence avec les connaissances actuelles** (sur l'histoire naturelle de la maladie, la physiopathologie, la biologie)

– **Analogie avec d'autres facteurs de risque démontrés** : une hypothèse causale est renforcée par le fait qu'une relation causale analogue est connue, rendant l'hypothèse plus crédible ; *ex: l'hypothèse que l'inhalation d'amiante cause le cancer du poumon est rendue plus plausible par la connaissance qu'on a du rôle causal du tabac dans le cancer du poumon.*

La présence de ces trois derniers critères renforce l'hypothèse de causalité mais leur absence ne l'exclut pas

• **Conclusion : une seule étude épidémiologique ne permet pas d'affirmer un lien de cause à effet.**

OBJECTIF 18°: VERIFIER QUE LES RESULTATS OFFRENT UNE REPONSE A LA QUESTION ENONCEE ET OBJECTIF 19°: VERIFIER QUE LES CONCLUSIONS SONT JUSTIFIEES PAR LES RESULTATS.

Ces deux objectifs seront traités ensemble car ils sont très liés : en effet pour vérifier que les résultats offrent une réponse à la question posée, il ne suffit pas de vérifier l'adéquation entre l'objectif principal de l'étude et le résultat principal de l'étude mais également que la méthodologie employée permet bien de répondre à la question posée ; ceci revient à évaluer la validité interne de l'étude. Cette évaluation sert aussi l'objectif 19 car les auteurs peuvent avoir tendance à sous évaluer les biais.

1. LES QUESTIONS A SE POSER

- a. Les auteurs ont-ils répondu à la question posée (objectif) dans l'introduction ?
- b. Ont-ils donné une explication rationnelle des résultats observés et de leur interprétation ?
- c. Les auteurs ont-ils évalué la validité interne de leur travail, c'est-à-dire la fiabilité de leurs résultats? En d'autres termes leurs résultats sont-ils non biaisés et suffisamment puissants?
- d. Les auteurs ont-ils discuté le caractère généralisable (validité externe) de leurs résultats ?
- e. Dans un essai clinique, si la conclusion est basée sur les résultats d'une analyse intermédiaire,
 - La ou les analyse(s) intermédiaire(s) ont-elles été prévues dans le protocole ?
 - A t-on tenu compte de ces analyses intermédiaires dans le calcul du NSN ?
 - A t-on précisé un seuil de signification pour chacune d'elles pour fonder les conclusions statistiques ?
(cf objectif 9)
- f. Si des analyses en sous-groupes ont été effectuées, étaient-elles prévues dans le protocole, argumentées, avec un but précisé (exploratoire, OU confirmatoire)? La méthode utilisée permet-elle de fournir une conclusion valide pour ces analyses ? (La randomisation a-t-elle été stratifiée ? L'interaction a-t-elle été testée ? A t-on corrigé le risque alpha ? L'interprétation repose-t-elle sur l'interaction ?) (cf objectif 9)

2. OU CHERCHER LES INFORMATIONS

Dans toutes les sections : introduction pour vérifier l'objectif, méthodes & résultats pour évaluer la validité interne et la représentativité, discussion pour évaluer l'interprétation que les auteurs ont faite de leur travail.

3. RAPPELS THEORIQUES

1. VALIDITE INTERNE : elle permet de s'assurer que le résultat obtenu n'est pas dû à un biais, au hasard ou à un manque de puissance donc que la méthode utilisée pour l'obtenir est adéquate pour répondre à la question.

- Absence de biais : cf objectif 14.
- Réalité statistique du résultat : cf objectif 16.
- Calcul du nombre de sujets réalisé pour garantir une puissance suffisante

2. VALIDITE EXTERNE : elle regroupe les notions de cohérence externe et de représentativité.

- **Cohérence externe** : permet de s'assurer que le résultat obtenu n'est pas isolé mais s'intègre dans un contexte logique (cohérence avec les connaissances expérimentales, épidémiologiques, les études de même nature). Pour les enquêtes épidémiologiques observationnelles, il s'agit de s'assurer que les arguments en faveur d'une relation de cause à effet entre l'exposition et la maladie sont rassemblés.
- **Représentativité** : les résultats sont-ils extrapolables aux patients rencontrés couramment dans la pratique médicale ? En d'autres termes, les patients de l'étude sont-ils représentatifs des patients vus en pratique médicale courante (même définition de la maladie, pas de sélection excessive (attention aux critères d'inclusion trop « durs ») ? L'objectif 21 détaille ce point.

OBJECTIF 20 : INDIQUER LE NIVEAU DE PREUVE DE L'ETUDE (GRILLE DE L'HAS / ANAES)

1. LES QUESTIONS A SE POSER :

- Quel est l'objectif de l'étude?
- Quel est le type d'étude? Est-il le plus approprié pour répondre à l'objectif?
- Les biais sont-ils importants dans cette étude ?
- La puissance de l'étude est-elle satisfaisante ? L'effectif repose t-il sur un calcul du nombre de sujets nécessaire ? Quel est le taux de sorties d'étude ou de données manquantes pour l'analyse principale ?

2. OU CHERCHER LES INFORMATIONS ?

Dans toutes les sections : introduction (objectif); méthodes; résultats; discussion (limites de l'étude)

3. RAPPELS THEORIQUES

DEFINITION DU NIVEAU DE PREUVE (ANAES/HAS)(4)

Le niveau de preuve d'une étude caractérise la capacité de l'étude à répondre à la question posée. Cette capacité se juge, d'une part, par la correspondance de l'étude au cadre du travail (sujet, population, paramètres de jugement pris en compte), et d'autre part par les caractéristiques suivantes :

- l'adéquation du protocole d'étude à la question posée,
- l'existence ou non de biais importants dans la réalisation,
- l'adaptation de l'analyse statistique aux objectifs de l'étude,
- la puissance de l'étude et en particulier la taille de l'échantillon.

Une classification générale du niveau de preuve d'une étude peut être proposée à partir des classifications de la littérature et des composantes vues ci-dessus :

- un fort niveau de preuve** correspond à une étude dont :
 - ✓ le protocole est adapté pour répondre au mieux à la question posée,
 - ✓ la réalisation est effectuée sans biais majeur,
 - ✓ l'analyse statistique est adaptée aux objectifs,
 - ✓ la puissance est suffisante ;
- un niveau intermédiaire** est donné à une étude de protocole similaire, mais présentant une puissance nettement insuffisante (effectif insuffisant ou puissance a posteriori insuffisante) et/ou des anomalies mineures ;
- un faible niveau de preuve** peut être attribué aux autres études.

Des distinctions plus fines ont été proposées par certains auteurs. Elles ne concernent que les études thérapeutiques et ne sont pas utilisables pour d'autres types d'études (diagnostic, causalité, cohorte).

GRILLE DE L'HAS/ANAES

GRILLE A INTERPRETER DANS LE CAS OU ON RECHERCHE UN LIEN DE CAUSALITE ENTRE UNE ACTION ET UN EFFET

Niveau 1	Essais comparatifs randomisés de forte puissance	<i>Recommandation Preuve scientifique établie</i>
	Méta-analyse d'essais comparatifs de forte puissance	
	Analyse de décision basée sur des études bien menées	
Niveau 2	Essais comparatifs randomisés de faible puissance	<i>Recommandation Présomption scientifique</i>
	Etudes comparatives non randomisées bien menées	
	Etudes de cohorte	
Niveau 3	Etudes cas-témoins	<i>Recommandation Faible niveau de preuve scientifique</i>
Niveau 4	Etudes comparatives comportant des biais importants	
	Etudes rétrospectives	
	Séries de cas	
	Etudes épidémiologiques descriptives (transversale, longitudinale)	

OBJECTIF 21°: DISCUTER LA OU LES APPLICATIONS POTENTIELLES PROPOSEES PAR L'ETUDE

1. LES QUESTIONS GENERALES A SE POSER :

- a. Quel est le niveau de preuve ?
- b. A quels patients les résultats sont-ils réellement applicables ? (cf objectif 3)
- c. Tous les critères cliniquement importants ont-ils été pris en compte dans la conclusion ?
- d. La balance bénéfice-risque est-elle favorable ?
- e. Les applications proposées sont-elles une conséquence directe des objectifs de l'étude?
- f. L'étude permet-elle de faire cette recommandation?
- g.

2 . LES QUESTIONS A SE POSER

POUR UN ESSAI CLINIQUE

- La population de l'étude correspond-elle à la population habituellement traitée ?
Si non : existe t-il une cause de force majeure pour ne pas appliquer les résultats à mon patient ?
- Toutes les variables cliniquement pertinentes ont-elles été évaluées ? Le critère de jugement est-il clinique ? *Par exemple : A t-on dans un traitement pour le cancer pris en compte en plus de la mortalité, la qualité de vie ?*
- Les bénéfices dus au traitement dépassent-ils les risques et le coût encourus ?
- Si analyse en sous-groupe, **(cf objectif 9)**
 - ✓ cette analyse a-t-elle été prévue (visée confirmatoire, justification, la randomisation a-t-elle été stratifiée sur la variable définissant les sous-groupes (*exemple : sexe pour analyse chez les hommes, et les femmes*) et le risque alpha a-t-il été corrigé ?
 - ✓ l'effet est-il significatif et important ? (test de l'interaction)
 - ✓ l'effet est-il retrouvé dans d'autres études ?
- Si arrêt prématuré de l'essai, des règles d'arrêt avaient-elles été définies à l'avance dans le protocole ? Ces règles ont-elles été respectées ? **(cf objectif 9, analyse intermédiaire)**
- **L'essai est-il pragmatique ou explicatif? (cf addendum 2 : types d'études)**

POUR UNE EVALUATION DE PROCEDURE DIAGNOSTIQUE

- Dans le cadre qui nous intéresse, les résultats du test diagnostique sont-ils reproductibles et leur interprétation satisfaisante ?
 - ✓ A t-on des informations sur la reproductibilité du test diagnostique ? Qui étaient les évaluateurs ?
 - ✓ Les réponses « douteuses » lorsqu'elles existent ont-elles été prises en compte dans l'analyse ? classées dans quel groupe (positif ou négatif) ?
 - ✓ Un calcul NSN a-t-il été effectué pour garantir une précision correcte dans l'estimation des paramètres (cf les intervalles de confiance des Se, Sp, VPP, VPN.)
 - ✓ Le test diagnostique est-il utile compte tenu de sa valeur informative et de sa reproductibilité dans le cadre qui nous intéresse ?
- Les résultats du test sont-ils applicables à la population à laquelle on veut appliquer le test diagnostique ?
 - ✓ Les conditions d'application du test diagnostique sont-elles les mêmes que dans l'étude ?
 - ✓ La population cible vérifie t-elle les critères d'inclusion et ne viole t-elle pas les critères de non inclusion ?
- Les résultats du test changent-ils la prise en charge ? Le test est-il informatif : permet-il d'éliminer le diagnostic ou au contraire d'affirmer le diagnostic au vu du résultat et compte tenu de la probabilité pré-test ? quel est le « gain » diagnostique d'un test positif (probabilité post-test) ou quel est « la perte » diagnostique d'un test négatif (probabilité post-test) ?
- Les patients auxquels on veut appliquer le test seront-ils mieux après la conclusion du test ?
Les bénéfices dus au résultat du test dépassent-ils les risques et le coût encourus par le test diagnostique ?

POUR UNE ENQUETE EPIDEMIOLOGIQUE ANALYTIQUE OU A VISEE ETIOLOGIQUE

- La population à laquelle on veut extrapoler les résultats est-elle comparable aux sujets recrutés dans l'étude, pour l'exposition, la maladie, l'âge, l'ethnie, et les autres caractéristiques importantes ?
- Quel est le niveau de preuve de l'étude ?
- Le faisceau d'arguments en faveur d'une relation de cause à effet entre l'exposition et la maladie est-il rassemblé ?
- Le risque est-il élevé si l'exposition est poursuivie ?
- Quelles sont les conséquences si l'exposition est supprimée ou réduite ?

POUR UNE ETUDE PRONOSTIQUE

- Les patients de l'étude sont-ils similaires aux patients auxquels on veut extrapoler les résultats de l'étude ?
- la population de l'étude est-elle décrite avec précision ?
- Les résultats de l'étude impliquent-ils une prise en charge spécifique : un arrêt de traitement ou une mise sous traitement ?
- Les résultats de l'étude permettent-ils de rassurer ou de conseiller mes patients?
 - ✓ Le pronostic est-il valide (sans biais) ?
 - ✓ Le pronostic est-il précis ?
 - ✓ Le pronostic est-il généralisable ?

3. OU CHERCHER LES INFORMATIONS ?

Dans toutes les sections : introduction (objectif); méthodes; résultats; discussion (limites de l'étude)

4. NOTIONS THEORIQUES

4.1. ETAPES DU GUIDE DE LECTURE CRITIQUE D'UN ARTICLE

Lire de façon critique un article comporte 3 étapes

1. Les résultats de l'étude sont-ils valides (non biaisés) ?
2. Quels sont les résultats de l'étude ? Autrement dit, la taille de l'effet est-elle importante ?
3. Quelles sont les implications des résultats pour ma pratique ?

La question des applications potentielles des résultats d'une étude est la troisième étape qui touche la validité externe de l'étude.

Elle concerne l'application au niveau individuel, et au niveau de la population.

4.2. LES GRILLES DE LECTURE POUR CHAQUE TYPE D'ETUDE

Le groupe de l'Evidence Based Medicine (EBM) (5-11) a publié des grilles de lecture qui regroupent un ensemble de points à passer en revue afin d'analyser dans quelle mesure les résultats sont applicables à un type de patient donné.

On peut transposer ces grilles au niveau des populations.

Ces grilles sont détaillées dans les paragraphes 2, 3,4 et 5 de cet objectif 21.

OBJECTIF 22 : IDENTIFIER LA STRUCTURE IMRAD (INTRODUCTION, MATERIEL ET METHODE, RESULTATS, DISCUSSION) ET S'ASSURER QUE LES DIVERS CHAPITRES REpondent A LEURS OBJECTIFS RESPECTIFS.

1. LES QUESTIONS A SE POSER :

- L'article contient-il quatre chapitres (Introduction, Matériel et méthode, Résultats, Discussion)?
- Chaque chapitre répond-il à ses objectifs ? Quels éléments ont été omis ? Quels éléments ne devraient pas y figurer ?

2. OU CHERCHER LES INFORMATIONS ?

DANS TOUT L'ARTICLE, EN SUIVANT L'ORDRE DE L'ARTICLE.

3. RAPPELS THEORIQUES

A. STRUCTURE IMRAD

Section	Rôles et buts	Plan
I ntroduction	Raisons de l'étude	S anté publique C linique/physiopathologie L acunes dans les connaissances
	Enoncer la question étudiée	P opulation I ntervention/facteur étudié C omparateur O bservé (outcome)
M éthodes	Décrire ce qui a été fait Montrer que les résultats seront valides Permettre la réplication de l'étude	S chéma d'étude P opulation L ieu(x) I ntervention(s)/facteur étudié C ritère de jugement A nalyse
R ésultats	Donner les éléments de réponse à la question	Description de l'échantillon étudié et date. Résultats sur CJP Résultats sur CJS
D iscussion	Interpréter et critiquer les résultats Proposer une implication clinique	Résumé des principaux résultats Points forts et limites de l'étude Discussion de leur validité Généralisation et confrontation à la littérature Conclusion sur la question Implications de la conclusion

B. La partie Méthodes

La partie Méthodes est la partie de l'article détaillant quels "matériels & méthodes" ont été à la base des résultats de la publication et des conclusions.

Cette partie Méthodes suit, normalement, le plan suivant (SPLICA)

- **Schéma d'étude** : ses paramètres (variables selon le type d'étude).
- **Population étudiée** : critères d'inclusion, de non inclusion, nombre d'échantillons, mode de randomisation

- **Lieu de l'étude** : structure hospitalière ou structure ambulatoire, nombre de structures (étude multicentrique)
 - **Intervention(s) Facteur étudié**: caractéristiques des interventions (chirurgicales, médicamenteuses ...) et leur durée, intervalle éventuel entre elles si plusieurs
 - **Critère(s) d'évaluation** : bien noter le critère d'évaluation principal (qui doit avoir été la base du calcul du nombre de sujets nécessaires et dont les résultats sont la base de la conclusion). Mode de recueil et d'évaluation (Notion d'aveugle) ...Noter si ce critère a été choisi après collecte des données de l'étude et pourquoi.
 - **Analyse**: calcul du nombre de sujets nécessaire (vérifier si les paramètres de ce calcul sont précisés), plan d'analyse (par exemple : analyse en intention de traiter pour un essai clinique, analyses principales, secondaires, analyses intermédiaires, analyses en sous-groupes, principe de construction des modèles multivariés), outils statistiques utilisés, valeur du risque α choisi.
- A ces points s'ajoutent les considérations réglementaires, éthiques.

Remarque : Cette question concerne la forme (structure) de l'article. => **Ne pas discuter** le contenu de l'article, la pertinence des méthodes, leur validité

OBJECTIF 23 : FAIRE UNE ANALYSE CRITIQUE DE LA PRESENTATION DES REFERENCES

1. LES QUESTIONS A SE POSER :

- Quel type de classement a été utilisé ?
- Les références sont-elles présentées de façon complète ?

2. OU CHERCHER LES INFORMATIONS ?

A LA FIN DE L'ARTICLE ET DANS L'ARTICLE POUR LES APPELS DANS LE TEXTE.

3. RAPPELS THEORIQUES

A. FORME D'UNE REFERENCE

- Auteurs : nom, initiales du prénom. Si plus de 6 auteurs: mettre « et al. », séparés par des virgules. (Remarque : parfois « et al. » est mis après les 3 premiers auteurs)
- Titre de la publication
- Support :
 - Revue : titre abrégé international (Index Medicus) ex: N Engl J Med, JAMA
 - Livre : nom de l'éditeur et lieu de publication
- Année de publication
- Volume et numéro de publication
- Première et dernière pages

Remarque: Si l'article est dans un ouvrage collectif : références de l'article et références de l'ouvrage (auteurs et titre de l'ouvrage en plus du nom de l'éditeur et du lieu de publication)

EXEMPLES DE REFERENCES

- **Article de revue** : Santen SA, Holt DB, Kemp JD, Hemphill RR. Burnout in medical students: examining the prevalence and associated factors. *South Med J* 2010; 103(8): 758-63.
- **Ouvrage** : Caplan LR. *Posterior circulation disease: clinical findings, diagnosis and management*. Cambridge, Mass: Blackwell Science, 1996: 555-556
- **Article dans un ouvrage collectif** : Goadsby PJ, Silberstein, SD, eds. *Pathophysiology of migraine: a disease of the brain*. In: Goadsby PJ, et al. *Headache*. Boston, Mass: Butterworth-Heinemann; 1997: 5-25.
- **Rapport** : Stoebner A, Lehmann M, Sancho-Garnier H. *Evaluation d'une action de prévention du tabagisme en entreprise*. Rapport février 1996

B. CLASSEMENT DES REFERENCES :

- Auteur-année
 - Dans le texte : [nom auteur, année]
 - Liste à la fin de l'article: par ordre alphabétique selon le nom de famille du premier auteur
- Alphabétique-numérique
 - Dans le texte : [numéro]
 - Liste à la fin de l'article: numéro par ordre alphabétique selon le nom de famille du premier auteur
- Numérique-séquentiel (Vancouver)
 - Dans le texte : [numéro]
 - Liste à la fin de l'article: numéro par ordre d'apparition dans le texte

OBJECTIF 24 : FAIRE UNE ANALYSE CRITIQUE DU TITRE

1. LES QUESTIONS A SE POSER :

- Le titre est-il informatif ou indicatif ?
- Le titre est-il attractif ?

2. OU CHERCHER LES INFORMATIONS ?

- Dans le titre, l'introduction et les méthodes

3. RAPPELS THEORIQUES

A BUT DU TITRE : ANNONCER LE CONTENU SIGNIFICATIF DE L'ARTICLE

B. PRINCIPES :

- Concision
- Précision
- Mots les plus informatifs au début (voire à la fin)
- Éviter jargon et abréviations

C. Contenu : PICOS

- Population
- Intervention (Facteur étudié)
- Comparaison (le cas échéant)
- Outcome (ce qui est observé, résultat)
- Schéma d'étude

D. Types :

- • Indicatif : orientation générale du contenu de l'article
- • Informatif : fournit des éléments précis sur le contenu significatif de l'article

E. Forme :

- Syntaxe souple
- Attractif

F. Exemples

Facteurs de risque environnementaux (Facteurs étudiés) des Lymphomes Malins Non Hodgkiniens (Outcome):
une étude cas-témoins (Schéma d'étude) de population en Languedoc-Roussillon, France (Population/Lieu)
=>titre informatif

• Les troubles (Outcome) liés à l'usage de drogues illicites (Facteurs étudiés) chez les conducteurs (Population)
par comparaison avec les étudiants (Comparaison) = >titre indicatif car aucune des composantes n'est précise.

ADDENDUM 1°: BASES EPISTEMOLOGIQUES DE LA CONNAISSANCE BIOMEDICALE

En première approximation, on peut définir l'épistémologie comme " l'étude de la constitution des connaissances valables". C'est là l'objet de la recherche bio-médicale. L'épistémologie pose trois grandes questions:

- Qu'est ce que la connaissance (la question gnoséologique) ?
- Comment est-elle constituée ou engendrée (la question méthodologique) ?
- Comment apprécier sa valeur ou sa validité ?

Très schématiquement, la connaissance bio-médicale a pour but final la Santé Humaine, donc d'une part la compréhension des mécanismes, internes ou externes à l'homme, qui déterminent les maladies, leurs évolutions et leurs conséquences et d'autre part la détermination des moyens qui peuvent "soigner et prévenir". Cette connaissance vise donc, par essence, à prouver qu'un facteur "causal" a une conséquence sur un facteur "objet". Par exemple:

- le tabac (facteur causal ou de risque) peut-il induire un cancer du poumon (objet ou conséquence) ?,
- un traitement donné (facteur causal) peut-il arrêter l'évolution d'une pathologie infectieuse ?,
- la pratique d'un sport (facteur causal) peut-elle prévenir l'apparition de maladies cardiaque ?

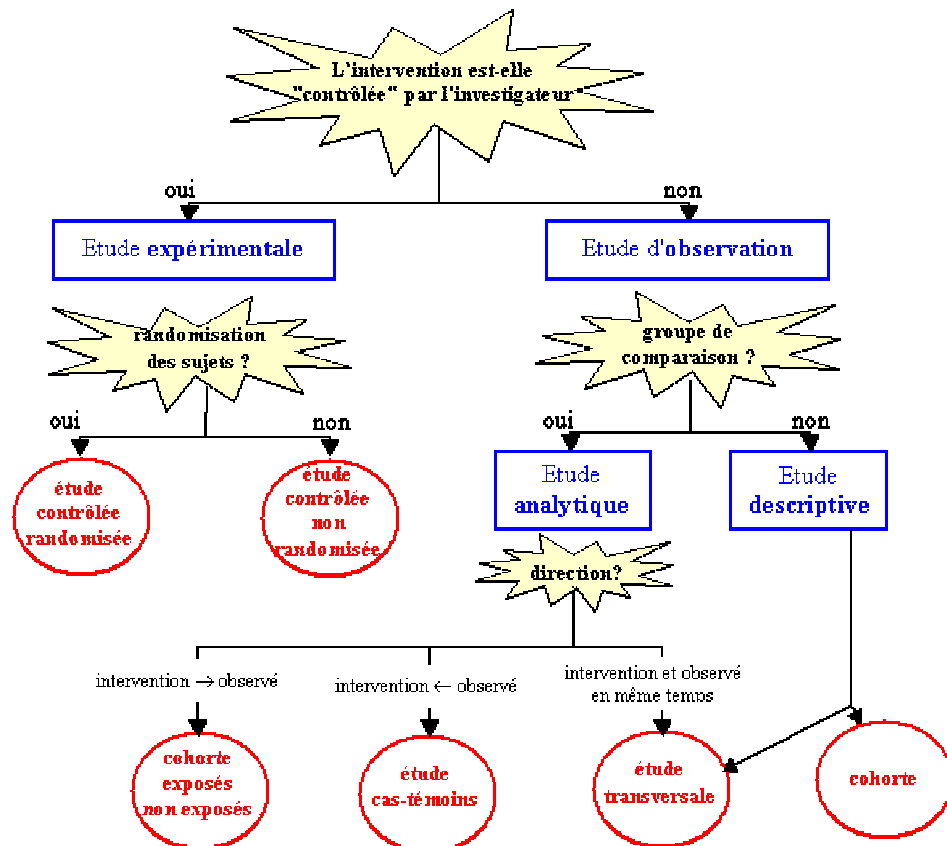
La méthode expérimentale: Quand on cherche une cause à quelque chose, la simple observation ne suffit pas. Il faut éliminer, avant d'observer, tous les facteurs existants qui pourraient induire cette chose (objet). Si le déclenchement de la cause induit toujours le même effet sur la chose, alors on en déduit le lien, ou lien causal, entre la cause et la chose. On a appliqué la méthode expérimentale : (1) bien définir la cause et l'objet, (2) éliminer tous les facteurs autres que la cause, (3) appliquer la cause et (4) observer la chose.

La causalité: Pour prouver un lien causal entre un facteur et un effet lorsque l'expérimentation n'est pas possible (champ de l'épidémiologie de population), il faut un faisceau d'arguments.

Sir Austin Bradford Hill a proposé en 1965 une série de critères pour évaluer les rapports épidémiologiques (le détail en est donné dans l'objectif 17).

- **En biologie et santé:** la notion de variabilité est systématiquement présente. Toute mesure sur ou à partir d'organismes vivants est entachée de variabilité inter-organisme et intra-organisme. La taille ou la glycémie à jeun des enfants de 12 ans et 3 mois ne sont pas identiques chez tous les individus. La glycémie à jeun chez le même individu n'est pas égale d'un jour à l'autre. Les causes d'un phénomène sont donc a priori multifactorielles. Face à cette multiplicité des causes possibles du même phénomène et alors que l'effet est "individu-dépendant",
- que signifie "prouver un lien causal" entre cause et objet pour une population d'individus ?
- La méthode expérimentale étant la méthode de choix, comment s'y prendre dans le cas où elle n'est pas applicable et notamment en Santé humaine où les questions éthiques sont omniprésentes ?

1. CLASSIFICATION DES TYPES D'ETUDES



2. ETUDE TRANSVERSALE

Synonymes: enquête de prévalence

Principe:

- "Photographie" d'un échantillon d'une population à un moment donné (souvent courte période, voire jour donné); pas de suivi des sujets, inclusion de tous les sujets de l'échantillon
- Recueil simultané d'informations sur le(s) problème(s) de santé, les caractéristiques et les déterminants / facteurs de risque éventuels présents

Avantages:

- Relativement simples et faciles à mener
- Comptabilisation du nombre de cas ⇒ **estimation de la prévalence**
- Génération d'hypothèses étiologiques
 - ✓ peuvent fournir des informations descriptives importantes sur l'étendue de l'exposition au facteur de risque et de la maladie dans une population
 - ✓ peuvent être utilisées comme première étape pour évaluer une relation possible entre exposition et maladie (à confirmer ou à réfuter dans d'autres études)

Inconvénients:

- Peuvent difficilement établir la séquence des événements entre exposition et maladie, à moins que l'exposition ne soit fixe (ex. groupe sanguin, sexe ou marqueurs génétiques); donc généralement peu utiles pour établir un lien de causalité entre une exposition et une maladie
- Ne sont généralement utiles que pour générer des hypothèses pour les études analytiques.
- Impossibilité d'évaluer l'incidence
- sujettes à de nombreux biais, y compris les problèmes liés au refus de participation, dont surtout bias de sélection: si échantillon mal représentatif de la population (impose, pour la sélection, l'utilisation de techniques de sondage)

Nota:

Peut être répétée et permettre ainsi de suivre l'évolution du problème au cours du temps.

3. ETUDE DE COHORTE

Principe:

- Etude d'une "tendance évolutive": enquête longitudinale
- Inclusion d'un échantillon de sujets, indemnes de la maladie ou du phénomène d'intérêt, sélectionnés selon des critères définis et à un moment donné et suivi dans le temps de ces sujets (souvent par examens périodiques)

Avantages:

- Comptabilisation du nombre de nouveaux cas \Rightarrow **estimation d'incidence**
- Etude simultanée de l'occurrence de plusieurs maladies
- Possibilité de suivre plusieurs facteurs de risque

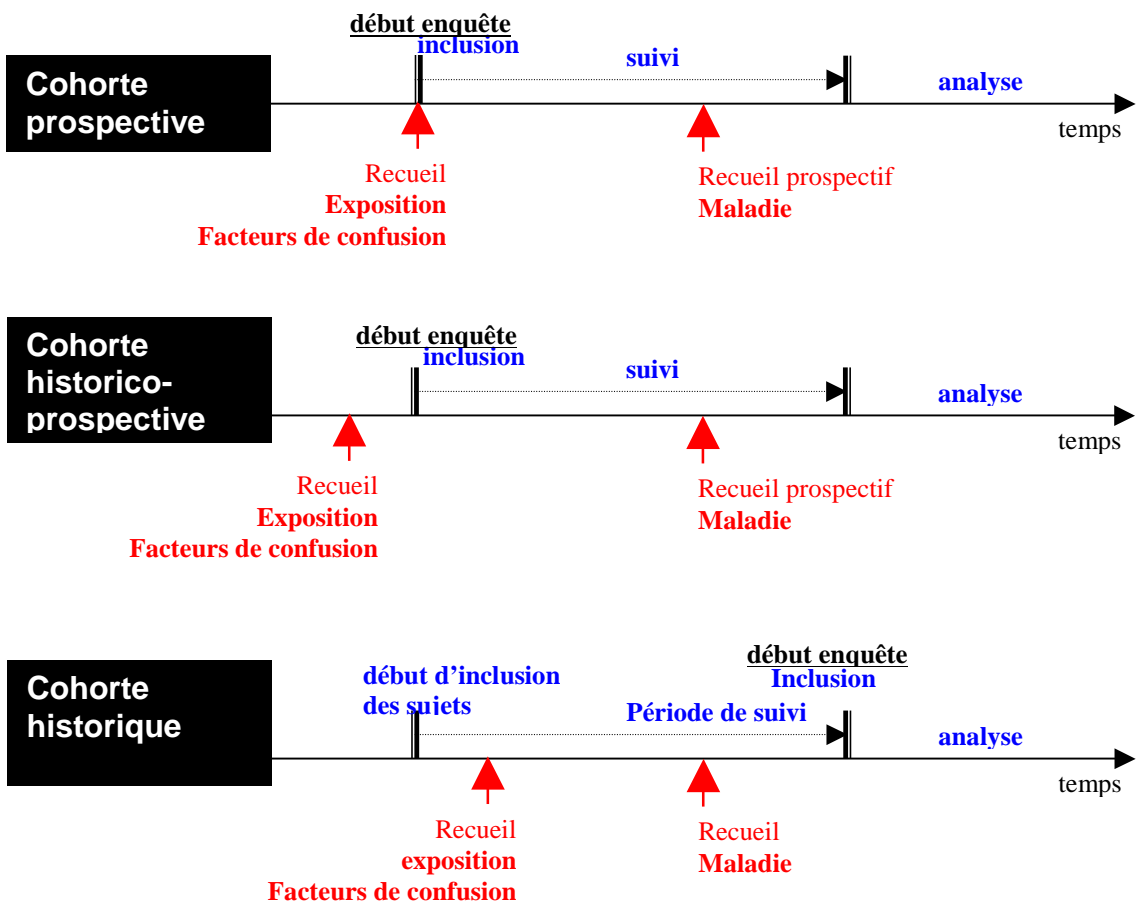
Inconvénients:

- Coût important
- Biais de classement (information) par modification dans le temps des critères diagnostiques
- Biais de sélection car possibilité de perdus de vue
- Risque de biais de confusion si l'objectif est de déterminer un lien entre une exposition et une maladie, car étude observationnelle. Les groupes analysés ne diffèrent pas uniquement par le facteur étudié. On ne contrôle pas les facteurs de confusion a priori.

4. ETUDE DE COHORTE EXPOSES - NON EXPOSES

Principe:

- Etude de cohorte dans laquelle les sujets, recrutés de la même manière, sont séparés en deux groupes , (1) exposé au facteur de risque et (2) non exposé, puis suivis pour dépister l'apparition de la maladie (ou du problème de santé) selon une procédure définie et identique pour les 2 groupes
- L' "intervention" est l'exposition au facteur de risque, l' "observé" la maladie.
- Etude épidémiologique, observationnelle, analytique
- Types de cohortes selon la période de recueil de l'exposition par rapport au début de l'enquête



Avantages:

- Détermination des taux d'incidence chez les exposés et les non exposés
- Permet de calculer le Risque Relatif et le Risque Attribuable dû au facteur
- Etude simultanée de l'occurrence de plusieurs maladies secondaires au même facteur de risque

Inconvénients:

- Non adaptée aux pathologies rares ou à délai long entre facteur de risque et maladie
- Coût important
- Biais de sélection
 - ✓ à l'inclusion, si pas bonne définition du facteur de risque
 - ✓ secondaire au suivi, si perdus de vue
- Biais de classement
 - ✓ si modification dans le temps des critères diagnostiques de la maladie
 - ✓ si changement de statut du sujet vis à vis de l'exposition
 - ✓ si recueil de la maladie non en insu de l'exposition
- Biais de confusion, si pas d'ajustement sur les facteurs de confusion

Nota:

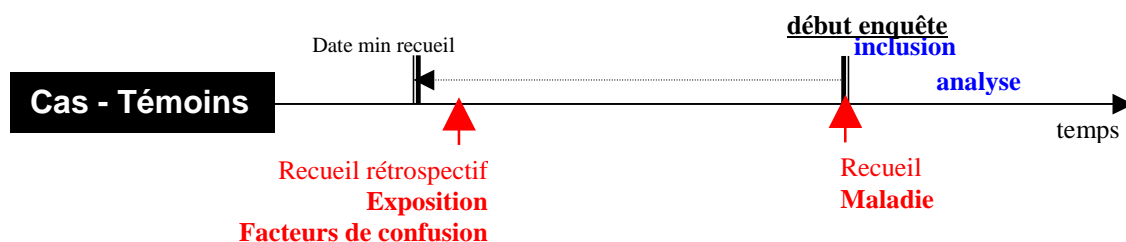
- Une cohorte historico-prospective ou historique ne peut se concevoir que si la reconstitution est exhaustive (tous les sujets et tous les paramètres: exposition, facteurs de confusion et maladie)
- 2 manières de constitution de la cohorte:
 - ✓ un échantillon de sujets éligibles puis classement dans un des 2 groupes suivant le statut exposé / non exposé
 - ✓ deux échantillons de sujets : un d'exposés et un de non exposés
- Le temps de suivi peut être variable d'un patient à l'autre (délai pour "voir apparaître" la maladie)
- Est recherché un "lien" entre facteur d'exposition et maladie. Or ce lien sera entaché par l'effet des facteurs de confusion. On effectuera donc toujours des analyses multivariées permettant d'ajuster sur les facteurs de confusion. **Le résultat à noter, dans l'analyse, sera donc toujours l'OR (Odds Ratio) ajusté ou le rapport de risques instantanés (RRI ou hazard ratio HR) et leur intervalle de confiance.**

5. ETUDE CAS - TEMOINS

Définition (glossaire) : Enquête rétrospective dans laquelle on interroge comparativement des malades (cas) et des non malades (témoins) sur leurs expositions dans le passé à des facteurs de risque. Les liens entre expositions et maladies (la mesure du risque de survenue de la maladie lié à l'exposition) sont résumés par des odds ratios et leurs intervalles de confiance.

Principe:

- Etude portant sur des sujets, recrutés et répartis en deux groupes selon leur statut vis à vis de la maladie (malades - ou cas - et non malades - ou témoins), et où on recherche une exposition à un facteur de risque dans les antécédents de ce sujets.
- L' "intervention" est l'exposition au facteur de risque, l' "observé" la maladie.
- Etude épidémiologique, observationnelle, analytique, rétrospective



Avantages:

- Permet d'estimer les taux d'exposition chez les cas et chez les témoins

- Permet d'avoir l'**Odds Ratio (OR)** - bonne approximation du risque relatif si la maladie est rare (prévalence faible) dans la population d'où est extrait l'échantillon; *attention: l'OR surestime toujours le lien*
- Souvent, c'est la seule possible pour les pathologies rares ou à délai long entre facteur de risque et maladie
- Durée d'étude et donc coût plus faible que pour les cohortes

Inconvénients:

- ne permet pas d'estimer les taux d'incidence de la maladie au sein des exposés et des non exposés
- ne permet d'obtenir ni le risque relatif ni le risque attribuable, ni la réduction absolue des risques
- **Biais de sélection**
 - ✓ si cas mal choisis (maladie latente). Pour limiter les biais, les critères diagnostiques doivent être objectifs, mesurables, reconnus et appliqués à tous les sujets.
- si témoins mal choisis. Pour limiter les biais, recruter les témoins sur la même période que les cas. Les témoins sont des sujets indemnes de la maladie et l'exposition du groupe témoin doit être représentative de celle des sujets indemnes de la population dont sont issus les cas.
- **Biais de classement**
 - ✓ de mémorisation, si exposition mal recueillie (dossier du patient, mémorisation du patient...)
 - ✓ exposition mieux recueillie par l'investigateur chez les cas que chez les témoins
 - ✓ Pour limiter ces biais, recueil des facteurs d'exposition en insu du statut (cas ou témoin) du sujet et outils de recueil de données (questionnaire) validés.
- Biais de confusion, si pas d'ajustement sur les facteurs de confusion.
 - ✓ Pour neutraliser les facteurs de confusion, on ajustera sur ces facteurs (cf ci dessous).

Nota:

- Le recrutement des cas peut se faire:
 - ✓ A partir des cas incidents: recrutement au fur et à mesure qu'ils se présentent
 - ✓ A partir des cas prévalents au début de l'enquête
 - ✓ Les cas incidents sont souvent plus représentatifs de l'ensemble des cas que les cas prévalents (cas prévalents ancienneté longue, facteurs personnels qui ont contribué à survie). Les cas incidents garantissent l'antériorité de l'exposition par rapport à la maladie. Avec les cas incidents, les biais de mémorisation concernant l'exposition sont moindres qu'avec les cas prévalents car les patients n'ont pas eu le temps de changer de mode de vie suite au diagnostic de la maladie.
- Le recrutement des témoins peut se faire:
 - ✓ A partir de sujets atteints d'une autre maladie (hôpital)
 - ✓ Au sein de la population générale
- Souvent choix de plus de témoins que de cas
- Les cas et les témoins doivent être aussi similaires que possible sauf pour l'exposition au facteur de risque. Très souvent lors de la planification, on apparie cas et témoins sur des facteurs de confusion connus (sexe, âge, catégorie socioprofessionnelle. L'appariement contribue à neutraliser ces facteurs de confusion. Toutefois il est nécessaire d'ajuster sur ces facteurs d'appariement.
- Ce qui est recherché est la preuve d'un "lien" entre facteur d'exposition et maladie. Or ce lien peut être entaché par l'effet des facteurs de confusion. On effectuera donc toujours des analyses multivariées permettant d'ajuster sur les facteurs de confusion. **Le résultat à noter, dans l'analyse, sera donc toujours l'OR (Odds Ratio) ajusté et son intervalle de confiance.**

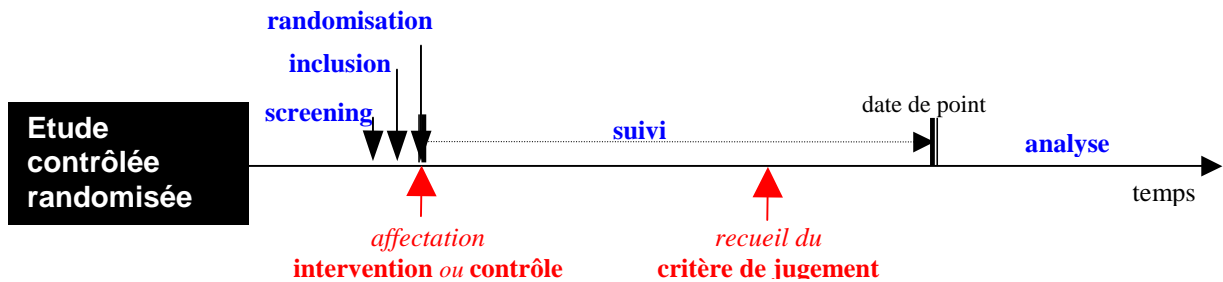
6. ETUDE CONTROLEE RANDOMISEE

Synonymes: essai clinique contrôlé randomisé quand les sujets sont des patients.

Cf glossaire: connaître toutes les définitions des expressions incluant le terme essai.

Principe:

- Etude expérimentale: les investigateurs décident du type et du déclenchement de l'intervention
- Contrôlée: existe un groupe contrôle recevant soit une autre intervention (souvent intervention de référence) soit aucune intervention (ou un placebo)
- Randomisée: affectation de la nouvelle intervention / intervention de comparaison (de référence, placebo) au sujet par tirage au sort



Avantages:

- Groupes (intervention et contrôle) sont créés par randomisation et donc comparables en tout sauf l'intervention. Donc, seule étude qui, si elle est bien conduite, évite tout biais de sélection et tout biais de confusion, connu ou inconnu.
- Absence de biais de classement si les critères de jugement sont précis, objectifs, fiables et valides, et recueillis de façon uniforme (standardisée) dans les 2 groupes, intervention et contrôle. Ce qui est fait par la procédure d'aveugle (ou insu) (sujet, investigateur, évaluateur).
- Puissance statistique satisfaisante (si peu de sorties d'étude, perdus de vue ...) car nombre de sujets nécessaire calculé pour détecter une différence cliniquement intéressante sur le critère de jugement principal, compte tenus des risques de 1^{ère} et de 2^{ème} espèce décidés au départ.

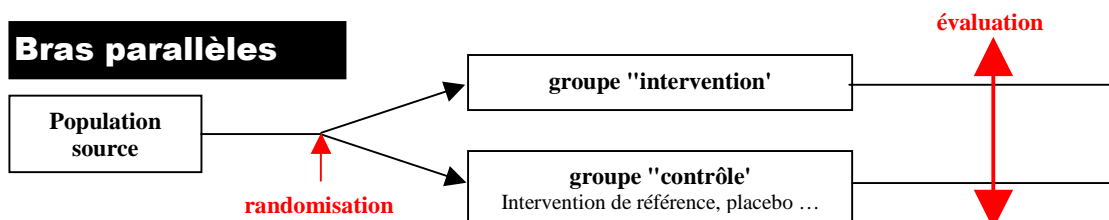
Inconvénients:

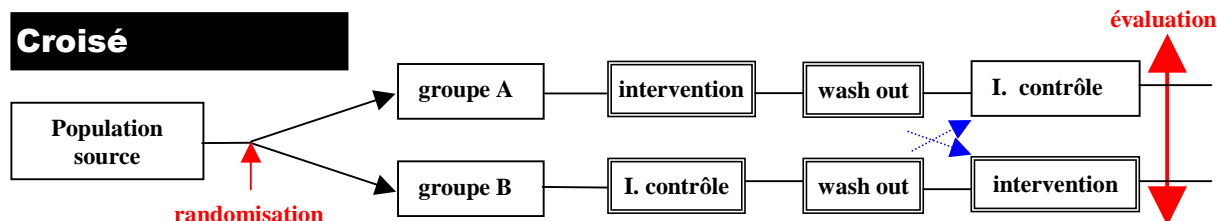
- Non possible dans certains cas (exposition à des toxiques, interventions nuisibles ...)
- Si sorties d'étude / perdus de vue en nombre trop important: possibilités de "détruire la randomisation" et de manque de puissance

Nota:

- Parfois plus d'un groupe contrôle
- Peut être multicentrique (plusieurs centres de recrutement). Risque alors de biais - de sélection / confusion si différences de procédure de screening et d'inclusion selon les centres. Le contrôle du biais peut être fait par randomisation stratifiée sur les centres puis, lors de l'analyse, ajustement sur les centres.
- Si des traitements concomitants sont donnés aux sujets, ces traitements doivent être identiques dans les deux groupes (sinon biais de confusion).
- Les 2 groupes étant comparables sur tout sauf l'affectation de l'intervention, le résultat à considérer est le **résultat brut (non ajusté) sur le critère de jugement**.
- Quelques particularités des essais d'intervention (*liste non exhaustive*)
 - risque de contamination entre les groupes surtout si suivi long
 - risque de biais de classement lié à l'évolution des techniques dans le temps
 - risque de biais lié à surveillance différentielle entre les groupes
 - risque de biais particuliers (ex : dépistage des cancers) :: lead-time, length time

Plans d'expérience





Les essais croisés nécessitent une pathologie chronique (retour à l'état initial après arrêt du traitement) et un critère de jugement mesurable à plusieurs reprises chez le même sujet.

Essai contrôlé randomisé de non infériorité

Principe: On concède a priori une perte d'efficacité au vu des avantages du nouveau traitement (meilleure tolérance, facilité d'utilisation plus grande, lourdeur du traitement plus faible, coût moindre). On teste l'hypothèse nulle suivante : le nouveau traitement a une efficacité inférieure à celle du traitement de référence.

Seuil de non infériorité : plus grande perte d'efficacité consentie par rapport au traitement de référence, étant donné les avantages que présente le nouveau traitement.

Implications : Il faut que les avantages du nouveau traitement confèrent une supériorité manifeste par rapport au traitement de référence, étant donné qu'on consent à une perte d'efficacité. => comparer le seuil de non infériorité et les avantages du traitement

Méthodologie pour éviter les biais

Essai contrôlé randomisé en double aveugle

Traitement de référence : efficacité pas altérée (administration correcte du traitement -dose, voie d'administration-, choix des patients, efficacité optimale, observance optimale).

Pas de traitement concomitant

CJP: clinique, objectif : choix du (des) seuil(s)

Validité de l'essai : interne avec 3ème bras placebo ou externe (confrontation essais antérieurs) : nouveau traitement doit être meilleur qu'un placebo

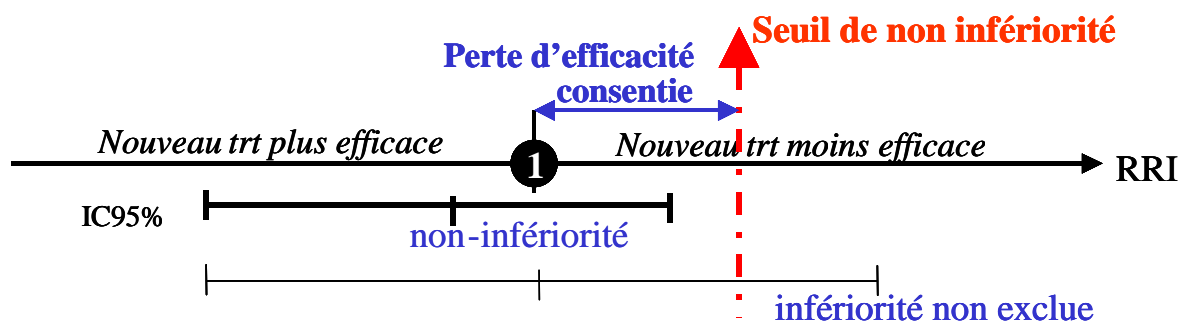
Minimiser les perdus de vue, écarts au protocole, et arrêts prématurés du traitement (<10%) qui tendent à diminuer la différence et à favoriser la non infériorité.

Stratégie d' analyse statistique : *per protocol* + ITT.

Conclusion fiable si les résultats des 2 analyses convergent.

On compare la borne *ad hoc* de l'intervalle de confiance à 95% au seuil de non infériorité.

Exemple1. critère de jugement principal : survenue de décès (RRI = rapport de risques instantanés)



7. ETUDE CONTROLEE NON RANDOMISEE

Synonymes: essai clinique contrôlé non randomisé quand les sujets sont des patients

Principe:

- Etude expérimentale: les investigateurs décident du type et du déclenchement de l'intervention
- Contrôlée: existe un groupe contrôle recevant soit une autre intervention (souvent intervention de référence) soit aucune intervention (ou un placebo)
- Allocation des patients non randomisée; ex de modes: allocation alternée des sujets ...

Avantages:

- Existence d'un groupe contrôle
- Contrôle de l'intervention, standardisation des procédures .

Inconvénients:

- Absence de randomisation donc risque de biais de sélection ++ ++ donc de non comparabilité des groupes au départ
-

Nota:

Le suivi et l'analyse sont faits comme si c'était une cohorte, le groupe "intervention" étant exposé et le groupe contrôle "non exposé". On parle de cohorte interventionnelle.

ADDENDUM 3°: EVALUATION D'UN TRAITEMENT

Particularités

Phases des essais médicamenteux

Phase 1	Volontaires, sains ou malades Etude de la tolérance (toxicité) Etude du mécanisme de traitement Recherche de la dose maximale tolérée
Phase 2	Patients malades Etude des conditions d'efficacité et des modalités d'administration Pharmacocinétique et dose minimale efficace Tolérance (effets secondaires fréquents) Efficacité (premières notions)
Phase 3	Patients malades Etude de l'efficacité et de la tolérance dans les indications du traitement Comparaison au traitement de référence / absence de traitement (placebo) Tolérance (effets secondaires fréquents)
Phase 4	Patients malades Etude de pharmacovigilance Effets secondaires rares / complications à long terme

L'analyse d'un essai de différence doit se faire en Intention de traiter (ITT) c'est à dire tous les patients sont analysés dans leur groupe initial de randomisation.

Modèle multivarié pour les études analytiques observationnelles ou notion d'ajustement sur des facteurs de confusion

ANNEXE 1 : OUTILS STATISTIQUES EN FONCTION DE LA PROBLEMATIQUE

A. VARIABLE A EXPLIQUER QUALITATIVE

Problématique	Outil statistique	Hypothèses
Estimation d'une fréquence théorique	$IC_{1-\alpha}$	NA
Comparaison d'une fréquence observée à une fréquence théorique	Test Chi2 de Pearson, test exact de Fisher, $IC_{1-\alpha}$	H0 : $F = F_{th}$ H1 : $F \neq F_{th}$
Comparaison de fréquences observées dans groupes indépendants	Test Chi2 de Pearson, test exact de Fisher	H0 : $F_A = F_B = F_C$ H1 : au moins une F est \neq
Comparaison de fréquences observées dans séries appariées	Test Chi2 de Mc Nemar	H0 : Egalité $F_{T1} = F_{T2}$ H0 : Différence $F_{T1} \neq F_{T2}$
Liaison de deux variables qualitatives	Test Chi2 de Pearson, test exact de Fisher, $IC_{1-\alpha}$ de OR	H0 : Indépendance, OR = 1 H1 : Liaison, OR \neq 1
Concordance (accord)	Indice kappa	H0 : pas d'accord au delà du hasard, kappa = 0 H1 : accord au delà du hasard, kappa \neq 0
Liaison var à expliquer qualitative sur séries indépendantes * var explicatives	Régression logistique multivariée (\Rightarrow OR ajusté)	H0 : indépendance; OR = 1 H1 : liaison; OR \neq 1
Liaison var à expliquer qualitative sur séries appariées * var explicatives <i>Ex: cas-témoins appariées</i>	Régression logistique conditionnelle (\Rightarrow OR ajusté)	H0 : indépendance; OR = 1 H1 : liaison; OR \neq 1

$IC_{1-\alpha}$ = Intervalle de confiance au niveau de confiance $1-\alpha$

NA : non applicable

B. VARIABLE A EXPLIQUER CENSUREE

Problématique	Outil statistique	Hypothèses
Estimation d'un taux de survie	Courbe de Kaplan Meier, courbe actuarielle	NA
Comparaison de k courbes de survie	Test du Logrank (k-1 ddl)	H0 : égalité des distributions de survie H1 : au moins une courbe de survie est \neq
Liaison Y censurée * var explicatives qualitatives ou quantitatives	Modèle multivarié de Cox (semi-paramétrique, à risques proportionnels) (\Rightarrow RRI ajusté)	H0 : indépendance; RRI = 1 H1 : liaison; RRI \neq 1

RRI = rapport de risques instantanés (en anglais HR: hazard ratio)

C. VARIABLE A EXPLIQUER QUANTITATIVE

Problématique	Outil statistique	Hypothèses
Estimation d'une moyenne théorique	$IC_{1-\alpha}$	NA
Comparaison d'une moyenne observée à une moyenne théorique	Test Student, $IC_{1-\alpha}$	H0 : $\mu = \mu_{th}$ H1 : $\mu \neq \mu_{th}$
Comparaison de 2 moyennes observées dans 2 groupes indépendants	Test Student, <i>Mann-Whitney-Wilcoxon</i>	H0 : $\mu_A = \mu_B$ H1 : $\mu_A \neq \mu_B$
Comparaison de 2 moyennes observées dans 2 séries appariées	Test Student apparié, <i>Wilcoxon apparié</i>	H0 : Egalité $\mu_{T1} = \mu_{T2}$ H0 : Différence $\mu_{T1} \neq \mu_{T2}$
Liaison de deux variables quantitatives	Coefficient de corrélation linéaire ρ de Pearson, ρ de Spearman	H0 : Indépendance, $\rho = 0$ H1 : Liaison, $\rho \neq 0$
Concordance (accord)	Coefficient de corrélation intraclasse,	H0 : pas d'accord, ICC= 0 H1 : accord, ICC $\neq 0$
Concordance (accord)	Graphe de Bland et Altman	
Liaison var à expliquer (Y) quantitative * 1 var explicative X qualitative	ANOVA, <i>test Kruskal-Wallis</i>	H0 : indépendance, $\mu_A = \mu_B = \mu_C$ H1 : liaison, au moins une μ est \neq
Liaison Y quant * X_i quantitatives	Régression linéaire multiple	H0 : indépendance, $\beta = 0$ H1 : liaison, $\beta \neq 0$
Liaison Y quant * X_i qualitatives et X_i quantitatives	Analyse de covariance	H0 : indépendance, $\beta = 0$ H1 : liaison, $\beta \neq 0$
Liaison Y quant à mesures répétées* X_i	Modèle mixte	H0 : indépendance, $\beta = 0$ H1 : liaison, $\beta \neq 0$
Capacité d'un test diagnostique à discriminer malades des non malades	$IC_{1-\alpha}$ de l'aire sous la courbe ROC (AUC)	H0 : AUC = 0,5 H1 : AUC $\neq 0,5$

Les outils statistiques en italiques sont des tests non paramétriques.

REFERENCES BIBLIOGRAPHIQUES

1. Mitchell LB, Exner DV, Wyse DG, Connolly CJ, Prystai GD, Bayes AJ, et al. Prophylactic Oral Amiodarone for the Prevention of Arrhythmias that Begin Early After Revascularization, Valve Replacement, or Repair: PAPA Bear: a randomized controlled trial. *Jama*; 2005. p. 3093-100.
2. Debrock C, Menetrey C, Bonavent M, Antonini MT, Preux PM, Bonnaud F, et al. [Prevalence of exercise-induced asthma in school children]. *Rev Epidemiol Sante Publique*; 2002. p. 519-29.
3. Hoffmann MH, Shi H, Schmitz BL, Schmid FT, Lieberknecht M, Schulze R, et al. Noninvasive coronary angiography with multislice computed tomography. *Jama*. [DIAGNOSTIC]. 2005 May 25;293(20):2471-8.
4. ANAES. Guide d'analyse de la littérature et gradation des recommandations. In: Santé ANdAedEeSHAe, editor.; 2000.
5. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*. 1993 Dec 1;270(21):2598-601.
6. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA*. 1994 Jan 5;271(1):59-63.
7. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*. 1994 Feb 2;271(5):389-91.
8. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*. 1994 Mar 2;271(9):703-7.
9. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. *JAMA*. 1994 Jul 20;272(3):234-7.
10. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. *JAMA*. 1994 May 25;271(20):1615-9.
11. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *JAMA*. 1993 Nov 3;270(17):2093-5.
12. Beuscart R., Bénichou J., Roy P., Quantin C. : Biostatistique: Montreuil, Omniscience, 2009.