



HAUTE AUTORITÉ DE SANTÉ

GUIDE METHODOLOGIQUE

Choix méthodologiques pour le développement clinique des dispositifs médicaux

Rapport d'évaluation

4 octobre 2013

L'argumentaire scientifique de cette évaluation sera téléchargeable sur
www.has-sante.fr

Haute Autorité de Santé
Service documentation et information des publics
2 avenue du Stade de France - F 93218 Saint-Denis La Plaine CEDEX
Tél. :+33 (0)1 55 93 70 00 - Fax :+33 (0)1 55 93 74 00

Avant Propos

Compte tenu des insuffisances fréquemment constatées dans l'évaluation de l'efficacité des traitements non pharmacologiques, la Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (CNEDiMTS) a souhaité identifier les méthodes et conditions permettant une évaluation clinique de qualité, notamment lorsqu'un essai contrôlé randomisé conventionnel ne peut être réalisé.

Ces méthodes ont toutefois leurs propres limites et doivent être réservées aux situations particulières où la réalisation d'un essai contrôlé randomisé conventionnel est impossible. Le choix de ce type d'étude devra être justifié scientifiquement. Au demeurant, l'essai contrôlé randomisé reste la référence pour toute étude clinique comparant l'efficacité de plusieurs traitements ou stratégies thérapeutiques.

La CNEDiMTS destine ce document aux industriels, aux structures de recherche et aux porteurs de projets. Il a pour objet de faire le point sur les méthodes comparatives pouvant être utilisées pour évaluer l'éventuel bénéfice clinique d'un nouveau dispositif médical ou d'une nouvelle technologie de santé, et de décrire les schémas d'étude possibles. Plus précisément, les objectifs étaient de :

- répertorier l'ensemble des méthodes potentiellement disponibles lorsque la randomisation et/ou l'aveugle sont impossibles à mettre en œuvre ;
- préciser leurs limites ;
- dans la mesure du possible, proposer des exemples d'application.

Sommaire

1. Introduction	7
2. Objectif	7
3. Avertissement	7
4. Contexte	8
4.1. Particularités de l'évaluation des DM et des technologies de santé	8
4.2. Phases du développement clinique	8
5. Méthode d'élaboration	10
5.1. Données bibliographiques	10
5.1.1. Recherche documentaire	10
5.1.2. Résultats de la recherche	10
6. Les difficultés méthodologiques	15
6.1. Choix du moment de l'évaluation	15
6.2. Population éligible et recrutement	15
6.3. Acceptabilité	15
6.4. Randomisation	16
6.5. Aveugle	17
6.6. Choix du groupe de référence ou comparateur	18
6.7. Facteurs liés à l'expérience des opérateurs	20
6.7.1. Courbe d'apprentissage	20
6.7.2. Volume des actes	20
6.8. Facteurs d'ajustement	21
6.9. Type d'analyse	21
6.9.1. Analyse en intention de traiter	21
6.9.2. Analyse en per-protocole	21
7. Méthodes permettant de pallier les difficultés identifiées	22
7.1. Obtenir l'aveugle ou compenser son absence	22
7.2. Autres type d'études contrôlées randomisées	24
7.2.1. Plans expérimentaux	24
7.2.1.1. « Zelen » ou « <i>randomised consent design trial</i> »	24
7.2.1.2. Essais basés sur l'expertise	28
7.2.1.3. « <i>Tracker trial design</i> » ou essais suiveurs	32
7.2.1.4. Essais en « <i>cluster</i> »	33
7.2.1.5. Essais en « <i>cross over</i> »	36
7.2.2. Analyses intermédiaires	38
7.2.2.1. Essais séquentiels	38
7.2.2.2. Essais avec randomisation adaptative	43
7.2.2.3. Méthodes bayésiennes	45
7.3. Etudes observationnelles comparatives non randomisées	47
7.3.1. Type d'étude	47
7.3.1.1. Etude de cohorte complète ou essai basé sur la préférence des patients	47
7.3.1.2. Etude observationnelle comparative prospective	48
7.3.2. Stratégies d'analyse	50
7.3.2.1. Score de propension	50
7.3.2.2. Variables instrumentales	52
8. En pratique	54
Annexe 1 : Glossaire	61
Annexe 2 : Tableaux récapitulatifs	63

L'équipe

Ce dossier a été réalisé par Isabelle FOURNEL (chargée de projet, Assistant Hospitalo-Universitaire, Dijon, isabelle.fournel@chu-dijon.fr) et Michel VANEAU (chef de projet, service évaluation des dispositifs, tél. : 01 55 93 37 56 e-mail : m.vaneau@has-sante.fr).

La recherche et la gestion documentaire ont été effectuées par Aurélien DANCOISNE (documentaliste, service documentation et information des publics, tél. : 01 55 93 73 39, e-mail : a.dancoisne@has-sante.fr) et Laurence FRIGERE (assistante documentaliste, service documentation et information des publics, tél. : 01 53 93 73 32, e-mail : l.frigere@has-sante.fr).

L'organisation des réunions et le travail de secrétariat ont été réalisés par Sandrine BOUVET et Fadela CHEBILI (tél. : 01 55 93 37 43, e-mail : s.bouvet@has-sante.fr et f.chebili@has-sante.fr).

Responsables hiérarchiques :

Hubert GALMICHE (Adjoint au chef de service évaluation des dispositifs, tél. : 01 55 93 37 48, e-mail : h.galmiche@has-sante.fr).

Catherine DENIS (Chef du service évaluation des dispositifs, tél. : 01 55 93 37 40, e-mail : c.denis@has-sante.fr).

Frédérique PAGÈS (Chef du service documentation et information des publics, tél. : 01 55 93 73 23, e-mail : f.pages@has-sante.fr).

Groupe de travail

Le groupe de travail était composé des professionnels suivants :

- Alain BERNARD
- Michel CUCHERAT
- Isabelle FOURNEL
- Hubert GALMICHE
- Pascal GIRAUX
- Bernard GUILLOT
- Jacques MACHECOURT
- Patrick MAISON

Groupe de relecture

Le groupe de relecture était composé des professionnels suivants :

- Michèle MORIN-SUROCCA
- Patrice NONY
- Françoise ROUDOT-THORAVAL
- Sophie STAMENKOVIC

Liste des abréviations

CNEDiMETS	Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé
DM	Dispositif Médical
HAS	Haute autorité de santé
IC	Intervalle de confiance
ITT	Intention de traiter
LPPR	Liste des produits et prestations remboursables
PP	Per protocol
SED	Service évaluation des dispositifs

1. Introduction

L'évaluation de l'efficacité d'un nouveau traitement doit reposer sur des études cliniques de qualité. En effet seuls des essais cliniques méthodologiquement bien menés permettent de porter un jugement sur la causalité entre le traitement étudié et l'effet observé, et en conséquence de conclure formellement sur l'effet du traitement.

Dès lors qu'une évaluation clinique a pour but de comparer l'efficacité de deux traitements, l'essai contrôlé randomisé en double aveugle bien conduit est le type d'étude offrant le meilleur niveau de preuve, permettant d'étayer des décisions médicales (1) et d'évaluer l'efficacité de ces traitements (2).

Randomisation, aveugle, et analyse en intention de traiter¹ constituent des critères majeurs de qualité méthodologique.

Dans la plupart des cas, la réalisation d'un essai contrôlé randomisé est possible. Cependant, dans certaines situations bien particulières qu'il convient de justifier, la réalisation de telles études sur les dispositifs médicaux (DM) est réputée difficile pour des raisons liées :

- à l'impossibilité de randomiser,
- au choix du comparateur,
- à la difficulté de mettre en place une procédure de double insu ;
- à la non-acceptation du patient ou du praticien;
- aux critères d'évaluation subjectifs dans certains cas.

Les autres types d'essais et les études observationnelles sont à réserver aux situations dans lesquelles l'essai contrôlé randomisé conventionnel n'est pas possible ; la décision de les utiliser doit faire l'objet d'une argumentation.

2. Objectif

Ce travail s'est intéressé aux aspects de l'évaluation clinique de l'efficacité d'un nouveau dispositif médical ou d'une nouvelle technologie de santé au moment de son développement, à la suite des études de faisabilité.

Il a pour objet d'identifier les méthodes et conditions permettant une évaluation clinique de qualité pour un DM. Il est destiné aux industriels, aux structures de recherche et aux porteurs de projets.

Plus précisément, les objectifs étaient de :

- répertorier l'ensemble des méthodes potentiellement disponibles lorsque la randomisation et/ou l'aveugle sont impossibles à mettre en œuvre ;
- préciser leurs limites ;
- dans la mesure du possible, proposer des exemples d'application.

3. Avertissement

Le choix d'une méthodologie différente de celle de l'essai contrôlé randomisé doit toujours être argumenté.

Les méthodes d'évaluation clinique présentées dans ce document sont celles à envisager lorsque les efficacités respectives de deux traitements, ou d'un traitement et d'un placebo,

¹ Dans le cas d'une étude visant à démontrer la supériorité d'un traitement

doivent être comparées. Ces méthodes ne préjugent pas de la pertinence clinique des critères de jugement utilisés dans ce but. De plus elles ne peuvent s'appliquer que sur des données de qualité suffisante. Aussi les études à réaliser doivent bénéficier d'un ensemble de mesures destinées à s'en assurer (protocole précis déposé dans un registre et/ou publié avant le début de l'étude, sélection et formation des investigateurs, méthode de collecte des données adaptée, monitoring, contrôle qualité et demandes de correction, etc...).

4. Contexte

4.1. Particularités de l'évaluation des DM et des technologies de santé

Les traitements non pharmacologiques peuvent être thérapeutiques, diagnostiques ou de compensation du handicap et concernent les techniques chirurgicales, les thérapeutiques interventionnelles (comme l'angioplastie), l'utilisation de dispositifs médicaux (par exemple endoprothèses vasculaires, stimulateurs cardiaques implantables, ultra-sons, stimulations électriques, orthèses).

L'évaluation de ces traitements est considérée comme étant un enjeu important, mais elle est souvent insuffisante (3). Le manque de données de bonne qualité constitue un obstacle important pour les agences d'évaluation et les décideurs.

Le cycle de vie d'une technologie a été décrit en 5 stades dans la littérature : futur (non encore développé), émergeant (juste avant leur introduction en pratique clinique), nouveau (dans la phase d'adoption), accepté (utilisation générale) et obsolète (ne devrait plus être utilisé) (3).

D'autres ont défini les technologies émergentes comme non encore adoptées par le système de santé, et les nouvelles technologies comme celles en phase d'adoption, disponibles en utilisation clinique depuis peu de temps (4).

Il est nécessaire d'évaluer une technologie avant sa diffusion, à savoir dans sa phase d'émergence (3). Ceci suppose de disposer d'informations sur l'efficacité clinique et la tolérance de cette technologie. En effet, la preuve de l'intérêt clinique est indispensable d'un point de vue éthique.

Le processus d'évaluation idéale d'une technologie de santé comporte 3 phases principales (3) : une phase d'alerte, une phase d'évaluation et une phase de surveillance.

La première phase, d'alerte ou « horizon scanning », repose sur un système de veille permettant d'identifier des technologies médicales nouvelles et émergentes.

La seconde phase consiste en l'évaluation approfondie (tolérance, efficacité) des technologies émergentes.

Enfin, la dernière phase repose sur la surveillance et la réévaluation ponctuelle de l'utilisation d'une technologie en pratique.

4.2. Phases du développement clinique

L'évaluation clinique de l'efficacité d'un nouveau traitement, qui est l'objet de ce document, est réalisée suite à la phase pré-clinique et aux études de faisabilité. Le protocole d'étude et

le recueil des données cliniques de la première utilisation chez l'homme ont une importance considérable dans ce contexte.

Les études de faisabilité sont proposées immédiatement à la suite de la phase pré-clinique. Elles permettent de mettre au point la technique et de déterminer les critères d'efficacité appropriés. Selon le contexte, une ou plusieurs études sont nécessaires pour répondre à différentes questions. Dans la plupart des cas, le type d'étude adapté méthodologiquement est l'essai prospectif non comparatif.

Le choix du seuil d'efficacité est primordial. En effet, avant d'envisager des études incluant un grand nombre de patients, un porteur de projet ou un industriel devrait vérifier que son nouveau DM est prometteur en termes d'efficacité. La première étape consiste donc à choisir un critère d'efficacité à partir des données de la littérature ou en s'appuyant sur des avis d'experts et de déterminer les probabilités d'efficacité et d'inefficacité. Ces probabilités permettront de construire des règles d'arrêt ou de poursuite en incluant le moins de patients possible.

Les études de faisabilité sont utiles pour (5) :

- La sélection des patients bénéficiant du nouveau dispositif médical

Cette étape consiste à préciser les formes cliniques de la pathologie et les caractéristiques des patients susceptibles de bénéficier du DM.

- La mise au point de la technique

Un DM implantable nécessite la mise au point de la technique d'implantation en décrivant les différents temps opératoires, le plateau technique et également le personnel nécessaires. Cette mise au point ne peut se faire que dans le cadre d'une étude clinique. Bien que la technique d'implantation puisse continuer de s'améliorer après cette étape, ceci ne doit pas retarder la mise en place d'une étude clinique pour la démonstration du bénéfice clinique.

- La mesure de l'efficacité clinique

L'étude de faisabilité est un préalable pour construire l'hypothèse à tester et calculer le nombre de sujets nécessaires dans une étude comparative. A ce stade du développement, la détermination du critère de jugement principal permettra donc de mesurer l'efficacité clinique du DM. Le critère doit être clinique et pertinent comme, par exemple, la réduction de la mortalité ou d'une complication cliniquement mesurable. L'utilisation d'un critère de jugement intermédiaire devra avoir été préalablement justifiée et le critère devra avoir été validé.

- Les complications et les risques

A la différence des études pharmacologiques n'impliquant pas d'acte médical, deux types d'événements indésirables peuvent être rapportés : ceux liés directement au DM et ceux qui sont en rapport avec la technique d'implantation ou opératoire. Les études de faisabilité ont également pour objectif d'estimer les principales complications, qui seront documentées lors de toutes les étapes, pour établir le futur rapport bénéfice/risque.

Ces étapes sont indispensables en fournissant des informations essentielles pour la démonstration ultérieure de l'efficacité par l'intermédiaire des essais contrôlés randomisés.

5. Méthode d'élaboration

Conformément à la réglementation, les experts ayant contribué aux travaux de la HAS ont rempli une déclaration publique d'intérêts, dont l'objet est de renseigner la HAS sur les éventuels conflits d'intérêts que certains des membres du groupe pourraient présenter avec un fabricant. Ces déclarations d'intérêts ont été rendues publiques sur le site de la HAS.

5.1. Données bibliographiques

5.1.1. Recherche documentaire

Une recherche documentaire systématique a été réalisée de janvier 2000 à juillet 2012, par interrogation des bases de données bibliographiques médicales :

- Medline (National Library of Medicine, Etats-Unis);
- The Cochrane Library (Wiley Interscience, Etats-Unis) ;
- BDSP (Banque de Données en Santé Publique) ;
- Les registres nationaux ;
- Les sites internet publiant des recommandations, des rapports d'évaluation technologique ou économique ;
- Les sites Internet des sociétés savantes compétentes dans le domaine étudié ;
- Les sources spécialisées notamment en épidémiologie et économie.

La recherche a été limitée aux publications en langues anglaise et française. Une veille a été réalisée jusqu'au 27 juin 2013.

Les évaluations technologiques, les recommandations, les conférences de consensus, les méta-analyses, les revues systématiques, les essais contrôlés randomisés ou non, les études comparatives et les études de cohortes ont été recherchés.

La stratégie de recherche et la liste des sources interrogées sont détaillées dans le paragraphe suivant. Cette recherche a été complétée par la bibliographie des experts consultés et les données fabricants.

5.1.2. Résultats de la recherche

La stratégie de recherche dans la base de données bibliographiques Medline est construite en utilisant, pour chaque sujet, soit des termes issus de thésaurus (descripteurs du MESH pour Medline), soit des termes libres (du titre ou du résumé). Ils sont combinés avec les termes décrivant les types d'études.

Le *tableau 1* présente la stratégie de recherche dans la base de données Medline et reprend les étapes successives et souligne les résultats en terme de :

- **Nombre références uniques identifiées : 1172**
- **Nombres de références analysées : 206**
- **Nombre de références retenues : 96**

Les résultats de la veille sur Medline ne sont pas inclus dans le tableau.

Tableau 1 : Stratégie de recherche dans la base de données Medline

Type d'étude / sujet	Période de recherche	Nombre de référence
Termes utilisés		
Choix méthodologiques pour le développement clinique des dispositifs médicaux – Recherche étendue		
Recommandations	01/2005 – 08/2010	128
Etape 1 ET Etape 2 ET Etape 3	(design[ti] OR designing[ti] OR method[ti] OR methods[ti] OR model[ti] OR modeling[ti] OR modelling[ti] OR statement[ti] OR reporting[ti] OR report[ti] OR process[ti] OR processing[ti] OR nonpharmacologic[ti] OR blinding[ti] OR blind[ti] OR checklist[ti] OR elaboration[ti]) ("Epidemiologic Research Design"[Mesh] OR "treatment outcome"[Mesh] OR "Publishing/standards"[Mesh] OR "Bias (Epidemiology)"[mesh] OR "Research Design/standards"[Mesh] OR delphi technique[mesh]) (recommendation* OR guideline* OR statement* OR consensus OR position paper)/ti OR (Health Planning Guidelines)/de OR (Practice Guideline OR Guideline OR Consensus Development Conference OR Consensus Development Conference, NIH)/pt	
Meta-analyses, revues systématiques	01/2005 – 08/2010	256
Etape 1 à étape 2 ET Etape 4	(metaanalys* OR meta-analys* OR meta analysis OR systematic review* OR systematic overview* OR systematic literature review* OR systematical review* OR systematical overview* OR systematical literature review* OR systematic literature search)/ti OR Meta-Analysis/pt OR Cochrane Database Syst Rev/so	
Essais contrôlés randomisés	01/2005 – 08/2010	219
Etape 1 à étape 2 ET Etape 5	random*/ti OR (Random Allocation OR Double-Blind Method OR Single-Blind Method OR Cross-Over Studies)/de OR Randomized Controlled Trial/pt	
Essais contrôlés	01/2005 – 08/2010	33
Etape 1 à étape 2 ET Etape 6	random*/ti,ab OR (Random Allocation OR Double-Blind Method OR Single-Blind Method OR Cross-Over Studies)/de OR (Randomized Controlled Trial OR Controlled Clinical Trial OR Multicenter Study)/pt	
Etudes de cohortes	01/2005 – 08/2010	18
Etape 1 à étape 2 ET Etape 7	(cohort* OR longitudinal stud* OR follow-up stud* OR prospective stud* OR retrospective stud*)/ti OR (Cohort Studies OR Longitudinal Studies OR Follow-Up Studies OR Prospective Studies OR Retrospective Studies)/de	
Autres études	01/2005 – 08/2010	30
Etape 1 à étape 2 SANS Etape 8	Sans étapes précédentes	

Choix méthodologiques pour le développement clinique des dispositifs médicaux - Recherche initiale

Recommandations		01/2005 – 08/2010	6
Etape 1	(Bayes Theorem[majr] OR Research Design[majr] OR Models, Statistical[majr] OR Bayesian[ti] OR Bayes[ti] OR study design[ti] OR methodological[ti] OR statistic*[ti] OR strateg*[ti] OR Premarket[tiab]) AND (Clinical Trials as Topic[mesh] OR Comparative Effectiveness Research[mesh] OR Meta-Analysis as Topic[mesh] OR Practice Guidelines as Topic[mesh] OR Device Approval[mesh] OR Product Surveillance, Postmarketing[mesh] OR Reproducibility of Results[mesh] OR premarket[tiab] OR sample size[mesh]) AND (Equipment Safety[mesh] OR Equipment and Supplies[mesh] OR Technology, Medical[mesh] OR device*[ti])		
ET			
Etape 2	(recommendation* OR guideline* OR statement* OR consensus OR position paper)/ti OR (Health Planning Guidelines)/de OR (Practice Guideline OR Guideline OR Consensus Development Conference OR Consensus Development Conference, NIH)/pt		
Meta-analyses, revues systématiques		01/2005 – 08/2010	67
Etape 1			
ET			
Etape 3	(metaanalys* OR meta-analys* OR meta analysis OR systematic review* OR systematic overview* OR systematic literature review* OR systematical review* OR systematical overview* OR systematical literature review* OR systematic literature search)/ti OR Meta-Analysis/pt OR Cochrane Database Syst Rev/so		
Essais contrôlés randomisés		01/2005 – 08/2010	51
Etape 1			
ET			
Etape 4	random*/ti OR (Random Allocation OR Double-Blind Method OR Single-Blind Method OR Cross-Over Studies)/de OR Randomized Controlled Trial/pt		
Essais contrôlés		01/2005 – 08/2010	10
Etape 1			
ET			
Etape 5	random*/ti,ab OR (Random Allocation OR Double-Blind Method OR Single-Blind Method OR Cross-Over Studies)/de OR (Randomized Controlled Trial OR Controlled Clinical Trial OR Multicenter Study)/pt		
Autres études		01/2005 – 08/2010	141
Etape 1			
SANS			
Etape 8	Sans étapes précédentes		

Légende tableau

Mesh : Descripteur ; * : troncature ; de: descriptor, ti: title; ab: abstract; pt: publication type; so: journal title

• Sites consultés

Dernière consultation : juin 2013.

Information francophone :

- Agence d'Evaluation des Technologies et des Modes d'Intervention en Santé, Canada
- Agence Nationale de Sécurité Sanitaire (Anes), France
- Agence Nationale de Sécurité du Médicaments et des produits de santé (ANSM), France

- Association Française de Normalisation - AFNOR
- Bibliothèque Médicale AF Lemanissier, France
- Catalogue et Index des Sites Médicaux Francophones, France
- Centre fédéral d'expertise des soins de santé, Belgique
- Direction de la recherche, des études, de l'évaluation et des statistiques, France
- Evaluation des technologies de santé pour l'aide à la décision (Fédération hospitalière de France) – ETSAD, France
- Expertise collective de l'INSERM, France
- Haute Autorité de Santé, France
- Institut national de prévention et d'éducation pour la santé (INPES) France
- Institut de recherche et documentation en économie de la santé (IRDES), France
- Institut de la statistique et des études économiques (INSEE), France
- Institut de veille sanitaire (InVS), France
- La Documentation française, France
- Portail de la statistique publique française, France
- Société Française de Médecine Générale, France
- Unions Régionales des Caisses d'Assurance Maladie, France

Information anglophone :

- Adelaide Health Technology Assessment, Australie
- Agency for Healthcare Research and Quality, Etats-Unis
- Alberta Heritage Foundation for Medical Research, Canada
- Alberta Medical Association, Canada
- American College of Physicians, Etats-Unis
- American Academy of Orthopaedic Surgeons
- American Heart Association, Etats-Unis
- Australian Therapeutic Goods Administration
- Blue Cross Blue Shield Association, Etats-Unis
- BMJ Clinical Evidence, Royaume-Uni
- Canadian Agency for Drugs and Technologies in Health, Canada
- Canadian Task Force on Preventive Health Care, Canada
- Centers for Disease Control and Prevention Infection Control Guidelines, Etats-Unis
- Centre for Clinical Effectiveness, Australie
- Centre for Reviews and Dissemination, Royaume-Uni
- CMA Infobase, Canada
- CONSORT Group
- Guidelines and Protocols Advisory Committee, Canada
- Guidelines International Network
- Institute for Clinical Systems Improvement, Etats-Unis
- Minnesota Department of Health – health Technology Advisory Committee, Etats-Unis
- National Coordinating Centre for Health Technology Assessment, Royaume-Uni
- National Guidelines Clearinghouse, Etats-Unis
- National Health Services Scotland, Royaume-Uni
- National Institute for Health and Clinical Excellence, Royaume-Uni
- National institute for health research Horizon scanning centre, Australie
- National Institutes of Health, Etats-Unis
- National Library of Guidelines Specialist Library, Royaume-Uni
- New Zealand Guidelines Group, Nouvelle-Zélande
- New Zealand Health technology Assessment, Nouvelle-Zélande
- Ontario Medical Advisory Secretariat, Canada
- Regional Evaluation Panel, Royaume-Uni
- Scottish Intercollegiate Guidelines Network, Royaume-Uni
- Singapore Ministry of Health, Singapour
- U.S. Food and Drug Administration
- U.S. Preventive Services Task Force, Etats-Unis
- Veterans Affairs Technology Assessment Program, Etats-Unis

- **Veille**

Une veille a été réalisée jusqu'au 27 juin 2013 sur *Medline* sur la base des équations du tableau 1. Cette veille a permis de mettre en évidence 213 références supplémentaires.

Les sommaires des revues suivantes ont été examinés tout au long du projet : British Medical Journal (BMJ), Journal of the American Medical Association (JAMA), The Lancet, The New England Journal of Medicine; la presse quotidienne médicale, paramédicale et l'Agence Presse Médicale (APM).

6. Les difficultés méthodologiques

Certains principes méthodologiques inhérents aux essais randomisés portant sur les traitements pharmacologiques peuvent être plus difficilement applicables à l'évaluation des dispositifs médicaux et des technologies de santé.

Les problématiques liées à la mise en œuvre d'essais contrôlés randomisés en double aveugle dans le domaine du dispositif médical sont détaillés ci-dessous.

6.1. Choix du moment de l'évaluation

Le choix du moment le plus adapté de la vie d'un DM pour procéder à son évaluation clinique est un point à prendre en compte. Les DM font usuellement l'objet de modifications visant à les améliorer après leur lancement. Ainsi un essai trop précoce peut ne pas refléter les réelles performances des dispositifs médicaux ne prenant pas suffisamment en compte l'apprentissage de la technique. Dans ce cas, une évaluation défavorable peut davantage refléter une technique mal maîtrisée qu'une réelle inefficacité de la technique (6).

Au contraire, une évaluation trop tardive est responsable de l'utilisation de DM ou de technologies de santé sans aucune preuve de leur efficacité. Or il est nécessaire que l'évaluation ait lieu avant leur large diffusion. En effet, lorsque le DM ou la technologie sont largement répandus, il est difficile de faire adhérer les médecins à un protocole d'étude (3), car une technique déjà utilisée est souvent empiriquement considérée comme efficace.

Au cours du temps, les utilisateurs vont en modifier les applications. Ces évolutions peuvent invalider l'évaluation initiale (3).

6.2. Population éligible et recrutement

La faible taille de la population éligible constitue également une spécificité des études portant sur les dispositifs médicaux. En effet, la population concernée peut être beaucoup moins importante pour les dispositifs médicaux que pour la plupart des traitements médicamenteux (3), pouvant ne concerner que quelques centaines de patients dans certains cas (7). Dans ce cadre, un essai classique en groupes parallèles peut être plus difficile à mettre en place en raison de sa complexité et de son coût.

Le choix de la population étudiée est important (8). En effet, si les patients éligibles sont sélectionnés de manière trop stricte, le rapport bénéfice/risque pour le dispositif sera optimisé, mais la validité externe sera plus limitée. En revanche, une sélection plus large peut faciliter le recrutement et la possibilité de généralisation des résultats, mais risque d'échouer à cibler la population la plus susceptible de tirer bénéfice du nouveau traitement.

6.3. Acceptabilité

L'acceptabilité de l'étude pour les patients joue un rôle important dans l'évaluation des traitements.

L'obtention du consentement du patient est le pré-requis pour la réalisation d'essais cliniques. L'étape d'information des patients avant l'obtention du consentement doit permettre de leur donner une information claire documentée et loyale. En l'absence de

consentement des patients, la faisabilité de l'étude est remise en cause. Lorsqu'il existe des raisons de penser que le rapport bénéfice-risque entre les traitements est différent, les patients comme les chirurgiens peuvent préférer une intervention donnée, et refuser de participer à l'essai. Ceci est particulièrement vrai pour les nouvelles méthodes chirurgicales, qui peuvent concerner des situations d'urgence ou la pédiatrie.

Certains patients préfèrent choisir une intervention, et refusent la randomisation (6). Ces considérations peuvent gêner le recrutement des patients et rendre la randomisation difficile. La diffusion de la technique peut également être une source de difficultés pour convaincre les patients de participer à une étude clinique.

Des questions d'acceptabilité peuvent se poser également au chirurgien, s'il est intimement persuadé que la technique qu'il a l'habitude de pratiquer constitue la meilleure stratégie (9).

6.4. Randomisation

Les techniques de randomisation jugées adéquates sont l'utilisation de tables de nombres aléatoires, la génération informatisée du groupe de randomisation par exemple (10). Il est indispensable de procéder à une randomisation centralisée. L'utilisation d'enveloppes, y compris opaques et scellées, ne garantit pas l'imprévisibilité du traitement reçu dans le cadre des essais en ouvert. En effet celles-ci peuvent être ouvertes et selon le traitement proposé, le médecin peut refuser de l'appliquer à son patient par exemple. Ces techniques doivent être décrites dans le protocole.

Bien que capitale, la méthode de randomisation est décrite et adéquate dans à peine la moitié des essais non médicamenteux, avec une assignation secrète et adéquate dans moins d'un quart des cas, d'après une revue de la littérature portant sur la parution d'essais cliniques dans des journaux à impact factor élevé (11).

Plusieurs facteurs pourraient expliquer ces lacunes. D'une part, un essai randomisé est plus coûteux qu'une série de cas (12). D'autre part, la randomisation peut être jugée a priori impossible, le plus souvent pour des raisons pratiques (préférence pour un nouveau traitement potentiellement efficace...), ou plus rarement pour des raisons éthiques (3).

Parmi les raisons éthiques évoquées figure la comparaison d'un traitement invasif à un traitement non invasif, risquant d'entraîner un refus de la part des patients. Bien que souvent cité dans la littérature, cet argument est très discutable. En effet, la diffusion de traitements sans preuve de leur efficacité et sans évaluation du rapport bénéfice-risque est également une attitude non éthique.

De plus ces essais sont dans la plupart des cas possibles. En effet la littérature fournit de nombreux exemples d'évaluation de traitements invasifs versus non invasifs, comme par exemple la greffe de cellules souches dans la maladie de Parkinson (13), ou l'angioplastie coronaire versus traitement médical (14).

6.5. Aveugle

L'aveugle (également appelé insu) est un élément important des essais cliniques, car il permet de limiter les biais de classement ou de mesure liés à la subjectivité du médecin ou du patient. L'impossibilité de distinguer les traitements comparés est un élément crucial pour respecter l'aveugle. L'aveugle peut concerner le patient, le médecin administrant le traitement, le critère de jugement, et/ou le statisticien.

Une revue de la littérature publiée en 2003 a pu montrer que l'aveugle concernait à peine un quart des patients, 6% des médecins et 2/3 des évaluateurs dans les essais non pharmacologiques (11). Ceci est particulièrement dommageable, d'autant plus que l'influence du médecin était également bien plus marquée que dans les essais pharmacologiques (11). Or les essais en ouvert surestiment de 14% la taille de l'effet par rapport aux essais réalisés en double aveugle (15).

L'aveugle est plus souvent impossible pour les essais non pharmacologiques (16), pour des raisons éthiques ou pratiques (3). Dans ces situations, afin d'évaluer de la manière la plus objective possible l'efficacité d'un traitement non pharmacologique, des alternatives ont été développées.

Boutron et al ont synthétisé les différentes méthodes d'aveugle utilisées dans les essais non pharmacologiques publiés dans les revues avec un facteur d'impact élevé (17). L'aveugle peut être complet, partiel, ou ne porter que sur l'évaluation du critère de jugement.

Classiquement, l'aveugle concerne les participants ou personnels de santé ayant en charge le patient. Il s'agit de la définition usuelle de l'aveugle. On parle de double aveugle lorsque ni le patient, ni le clinicien ne connaissent le traitement reçu par le patient.

L'aveugle est complet lorsque l'intervention est simulée dans le groupe contrôle et en l'absence de contact entre l'opérateur (le plus souvent un chirurgien) et l'équipe assurant le suivi du patient. Malgré les difficultés pratiques pour les traitements interventionnels, certains essais ont été réalisés en aveugle complet, comprenant placebo identique, simulation de l'intervention, et absence de contact avec les chirurgiens.

Dans certains cas l'aveugle de ces personnes-clés est impossible (Tableau).

L'absence d'aveugle peut conduire à des biais tels les biais de sélection, de suivi, d'attrition ou de mesure (cf. annexe 1), remettant en cause la validité interne d'une étude (6).

Tableau 2 : Exemples d'impossibilité de réalisation de l'étude en aveugle d'après (18)

Aveugle patient	du Aveugle de l'opérateur	Aveugle de l'équipe assurant le suivi	Evaluation en aveugle
- Aspect ou perception du dispositif - Cicatrice qui révèle le type d'intervention (ex : chirurgie versus traitement non interventionnel, laparotomie versus coelioscopie...)	-Intervention chirurgicale -Aspect ou manipulation du dispositif	- Effets indésirables spécifiques et caractéristiques d'un des traitements administrés - Coloration ou traces caractéristiques laissés par le dispositif - Aspect radiologique évocateur d'un des traitements reçus - Cicatrice qui révèle le type d'intervention	- Organisation pratique : nombre excessif de visites ou de consultations, limitant l'acceptabilité pour le patient - Si consultation uniquement réalisée par une personne extérieure : le chirurgien est ramené uniquement à un rôle de technicien, ce qui pose un problème d'acceptabilité pour le chirurgien, et également pour le patient

6.6. Choix du groupe de référence ou comparateur

Le choix du groupe de référence ou comparateur est crucial dans les essais non pharmacologiques. Cette problématique ne concerne pas les essais dans lesquels le traitement évalué s'ajoute au traitement de référence², qui est utilisé seul dans le groupe contrôle.

Selon la 11^{ème} directive du Conseil des Organisations Internationales des Sciences Médicales concernant la recherche biomédicale sur des êtres humains, l'usage du placebo (ou d'un traitement inactif) peut être considéré comme éthiquement acceptable dans les circonstances suivantes :

- en l'absence de traitement ou de procédure efficace ;
- lorsque l'abstention d'un traitement ou d'une procédure avec une efficacité établie n'entraînera au pire qu'un inconfort temporaire ou un retard dans le soulagement des symptômes ;
- lorsque la comparaison avec un traitement ou une procédure efficace ne fournit pas de résultat scientifiquement fiable, et que l'administration d'un placebo n'ajoute pas de risque important de dommage irréversible (19).

Sur le plan éthique, il est difficile de proposer une intervention fictive invasive (20). En effet, plus la procédure est interventionnelle, plus il est difficile de justifier l'exposition des patients inclus dans un groupe témoin à des risques éventuellement importants, sans bénéfice attendu (21). Une objection importante à ce point de vue est que l'enjeu éthique est aussi celui des traitements reçus par les patients futurs, dans la mesure où la diffusion à grande échelle d'un traitement non évalué n'est pas éthique. Ceci suggère qu'il peut être important d'impliquer les patients inclus au sujet de l'intérêt global d'une étude.

En particulier, il a été avancé qu'il n'est pas éthique d'administrer un placebo à la place d'un traitement de référence dont l'efficacité serait démontrée (19), ou encore un « placebo invasif » non utile à la solidité de la démonstration. Ainsi, dans l'évaluation de la greffe de cellules souches en cas de maladie de Parkinson, le groupe témoin subissait une intervention fictive sous anesthésie générale, avec incision abrasion localisée de la corticale externe de la boîte crânienne (22). Le choix de cette chirurgie placebo a été largement controversé pour plusieurs raisons. D'une part, elle n'apporte aucun bénéfice aux patients, alors qu'elle comporte un risque (lié à l'anesthésie) (23), et d'autre part des alternatives existaient (24), comme l'absence de chirurgie ou la stimulation cérébrale profonde (25)

Une revue de la littérature a synthétisé les principaux placebos de chirurgie utilisés dans les essais non pharmacologiques (17). Pour la chirurgie et les interventions techniques, différentes méthodes sont rapportées, selon le temps de l'intervention. Ainsi, les patients peuvent être sous anesthésie générale, ou un champ opératoire peut masquer l'intervention. Dans certains cas, l'intervention a été simulée, en réalisant une incision similaire à celle du groupe traité, ou par l'injection d'un placebo.

En pratique, la chirurgie dite placebo est pratiquement impossible et restreinte aux cas où il n'y a aucun comparateur convenable, et lorsqu'elle comporte peu de risques (6). La standardisation des soins pré-opératoires (même position des patients ou du matériel), per-opératoires (durée de l'intervention, instruments, manipulation ou soins) et post-opératoires est également importante. Dans d'autres études, le chirurgien ayant fait le geste ne participe pas au suivi des patients.

² Essais dits « *on top* »

Boutron et al ont également rapporté les différents placebos possibles pour l'utilisation des dispositifs médicaux : prothèses placebo, dispositifs médicaux cachés; dispositifs médicaux identiques mais inactivés, ou activés mais rendus inefficaces, ou utilisation d'un appareil simulée (17,26) (Tableau 3 : Exemples de placebos).

Tableau 3 : Exemples de placebos utilisés en chirurgie et pour les dispositifs médicaux d'après (17)et (18)

		Placebo
Techniques chirurgicales	En période anesthésique	<ul style="list-style-type: none"> - Patients sous anesthésie générale - Patients masqués par un drap
	Pendant l'intervention	<ul style="list-style-type: none"> - Incisions cutanées pour obtenir une cicatrice similaire à celle de l'intervention testée (exemple : pontage aorto-coronariens avec une artère mammaire interne versus incisions cutanées seules chez les patients avec une angine de poitrine, chirurgie arthroscopique versus incisions cutanées seules chez des patients atteints d'arthrose) - Simulation de l'intervention (exemple : injection intracérébrale de cellules fœtales versus incisions cutanées et abrasion de la corticale externe de la boîte crânienne chez des patients atteints de la maladie de Parkinson)
	En post opératoire	<ul style="list-style-type: none"> - Cicatrices masquées - Application des mêmes pansements éventuellement imprégnés de sang sur des plaies réelles ou fictives pour avoir des aspects identiques dans une étude comparant la mini-cholécystectomie à la cholécystectomie coelioscopique, ou pansements similaires pour comparer exérèse du colon par laparotomie et colectomie coelioscopique) - Standardisation des soins post-opératoires / traitements associés
Dispositifs médicaux	DM placebo	<ul style="list-style-type: none"> - Placebo non strictement identique au traitement reçu - Prothèse similaire mais ne fournissant pas l'effet thérapeutique (par exemple sans production de chaleur) - Utilisation de DM différents cachés (exemple : sac de transfert pour aphérèse témoin, lunettes opaques pour le participant...)
	Aspect identique	<ul style="list-style-type: none"> - Utilisation d'un appareil identique inactivé (même durée, fréquence, précautions...) - Utilisation d'une machine identique activée : avec une barrière pour bloquer le traitement, en modifiant la position de la source du traitement pour qu'il n'y ait pas d'exposition au traitement

6.7. Facteurs liés à l'expérience des opérateurs

6.7.1. Courbe d'apprentissage

Une particularité des technologies de santé faisant appel aux dispositifs médicaux est l'impact de l'expertise de l'opérateur sur les performances de la technique (3). Des niveaux d'expérience différents peuvent conduire à des niveaux de performance différents lors de la réalisation des interventions. Un manque d'expérience peut influencer le résultat de l'essai en pénalisant la nouvelle thérapeutique testée (biais de performance) (27).

Ainsi la courbe d'apprentissage des opérateurs doit être prise en compte dans l'évaluation des techniques chirurgicales ou interventionnelles. Lors du développement d'un nouveau dispositif médical, des plans de formation et d'apprentissage doivent être prévus. En effet, les connaissances des chirurgiens et leur expertise sont autant de facteurs de variabilité des pratiques (6) qu'il est nécessaire de prendre en compte, par exemple par une période de rodage (28).

Une évaluation trop précoce risque de refléter les complications liées à l'apprentissage de la nouvelle intervention. Lors des interventions chirurgicales, l'impact de l'apprentissage peut être évalué par des indicateurs comme par exemple : la durée de l'acte ou le volume des pertes sanguines.

La variabilité de l'effet du traitement selon le niveau d'expertise de l'opérateur doit également être évaluée. L'importance de l'expérience de l'opérateur a été soulignée dans les essais comparant angioplastie et thrombolyse dans le traitement de l'infarctus du myocarde (29,30).

Ainsi l'évaluation d'une nouvelle technologie contre témoin risque d'être déséquilibrée en défaveur de la nouvelle technique, en raison de l'expérience de l'opérateur (6). L'évaluation doit donc intégrer l'effet de l'apprentissage, par exemple en enregistrant la formation et l'expérience (6). D'un point de vue pragmatique, cette phase d'apprentissage doit être prise en compte dans l'essai afin de pouvoir évaluer précisément le bénéfice apporté par le dispositif ou la technologie de santé.

6.7.2. Volume des actes

Le volume d'activité doit aussi être pris en compte dans l'évaluation d'une nouvelle technologie. Ainsi, près de 70% des études retrouvent une association significative entre résultats cliniques favorables et volume d'activité du médecin (31). Cette association est d'autant plus importante pour les interventions chirurgicales complexes, rarement effectuées, et à risque (20).

6.8. Facteurs d'ajustement

Il est parfois difficile d'isoler les effets des traitements non pharmacologiques de ceux des autres composants du système de soins. En effet, les DM sont souvent utilisés en conjonction avec d'autres interventions (chirurgie, diagnostique, surveillance...). L'évaluation de l'effet propre du DM sur le critère de jugement peut donc être difficile (3). Ainsi, les interventions chirurgicales dépendent non seulement du chirurgien (préférence, expérience...) mais également des autres membres de l'équipe (anesthésiste, infirmier), de la prise en charge pré et post-opératoire (6) et de l'organisation des soins.

Les pratiques pouvant varier dans les différents centres, il est indispensable de définir *a priori* très précisément dans le protocole les conditions de réalisation afin d'éviter des résultats biaisés liés aux types de pratiques. Cette standardisation doit être extrêmement détaillée, en décrivant précisément la procédure et les moyens nécessaires, qui doivent figurer dans le cahier des charges (8). Cependant, dans le cas où ces pratiques ne pourraient être reproductibles, on cherchera plutôt à recueillir les éventuelles différences entre les centres, afin de pouvoir ajuster l'analyse sur ces éléments.

6.9. Type d'analyse

6.9.1. Analyse en intention de traiter

L'analyse en intention de traiter consiste à analyser les patients dans leur groupe de randomisation, quel que soit le traitement reçu. Elle est à privilégier dans les essais de supériorité car il s'agit de l'attitude la plus conservatrice. Elle permet en effet de ne pas perdre le bénéfice de la randomisation et d'éviter les écarts au protocole qui ne seraient pas dus au hasard mais seraient en rapport avec le traitement administré.

6.9.2. Analyse en per-protocole

L'analyse en per-protocole consiste à analyser les patients selon le traitement qu'ils ont reçu, et non pas selon leur groupe de randomisation. Cette stratégie entraînant une augmentation de la différence de l'effet des stratégies comparées, elle est à privilégier dans les essais de non-infériorité ou dans les essais d'équivalence.

7. Méthodes permettant de pallier les difficultés identifiées

7.1. Obtenir l'aveugle ou compenser son absence

Lorsque l'aveugle du personnel soignant est impossible, une évaluation en aveugle du critère de jugement est envisagée. Cela permet de garantir une évaluation neutre de ce dernier. Dans ce cas, l'évaluation est réalisée par des évaluateurs indépendants de l'étude en aveugle du traitement reçu. L'évaluation peut également être centralisée pour des examens biologiques ou radiologiques, ou des extraits de l'examen clinique (vidéo, photos, enregistrement d'entretiens).

En revanche, l'aveugle de l'évaluation peut être pris en défaut :

- si les patients connaissent leur groupe de randomisation et doivent rencontrer l'évaluateur
- si le traitement est identifiable par l'évaluateur indépendant
- si en pratique on ne peut s'affranchir du suivi assuré par le chirurgien qui a pratiqué l'intervention
- si les documents destinés au comité d'évaluation sont envoyés par un investigateur qui a connaissance du traitement reçu

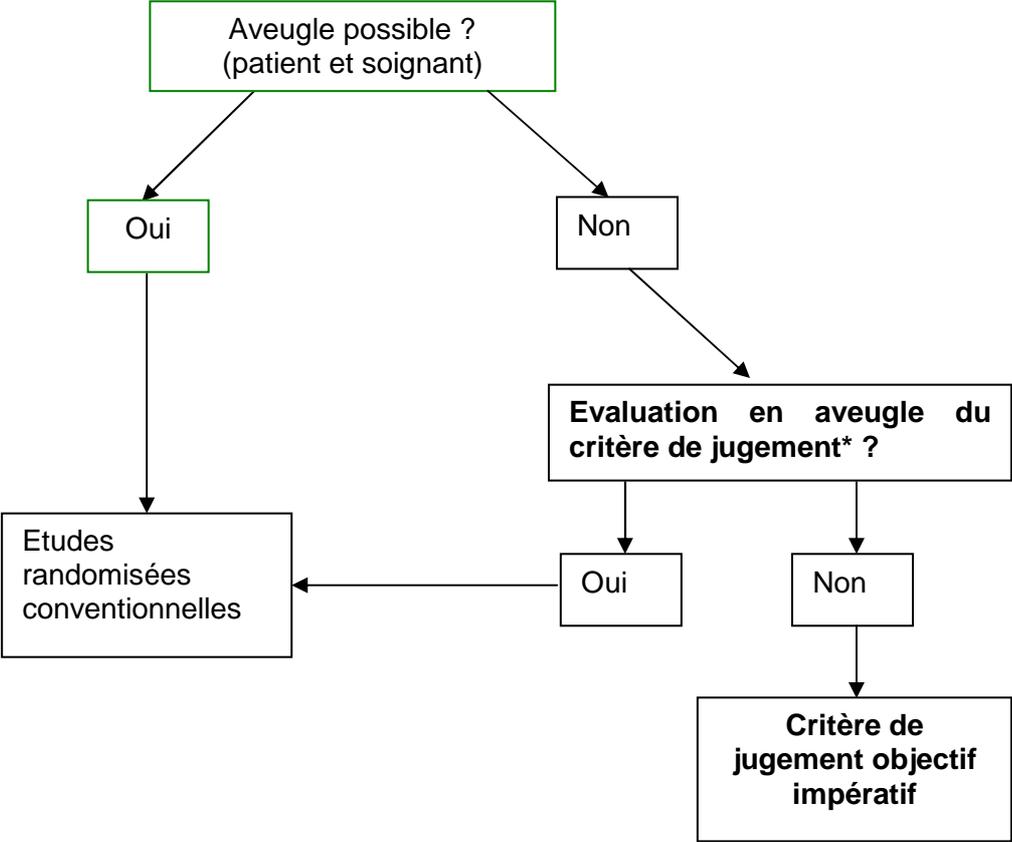
Dans les interventions chirurgicales, l'aveugle du patient peut être réalisé si le patient est sous anesthésie générale, ou masqué par un champ opératoire (17).

Dans certains cas, un comité d'adjudication indépendant des investigateurs est mis en place afin de contrôler le critère de jugement. L'aveugle des hypothèses de l'étude (ou aveugle partiel) est également une alternative décrite (17). Elle repose sur une information partielle délivrée aux participants, qui par exemple ne sont pas informés sur l'existence d'un traitement de référence, sur les hypothèses de l'étude (hypothèse de supériorité d'un traitement), ou sur la randomisation des patients selon un plan expérimental de Zelen modifié. Dans ce cas, pour des raisons éthiques, il est nécessaire d'informer le patient que, pour des raisons scientifiques, il ne sera pas informé de l'objectif spécifique de l'étude, mais qu'il aura accès à l'ensemble des informations à la fin de l'étude.

Lorsque l'aveugle ou ses alternatives sont impossibles, il est important de privilégier un critère de jugement le plus objectif possible, et de procéder à une évaluation en aveugle du critère de jugement (par exemple par un comité d'expert indépendant).

La **Figure 1** présente l'arbre décisionnel pour avoir le meilleur niveau de preuve vis-à-vis du critère de jugement.

Figure 1 : Arbre décisionnel pour limiter les biais de mesure dans les essais randomisés



* ou observateur indépendant ne connaissant pas le traitement

7.2. Autres type d'études contrôlées randomisées

L'essai en deux groupes (ou bras) parallèles³ doit être envisagé avant tout autre plan expérimental dès lors qu'il est nécessaire de conduire une évaluation clinique comparative. Dans ce cas le traitement étudié est comparé à un traitement contrôlé à l'aide de deux groupes de patients, constitués par randomisation de façon contemporaine et suivis en parallèle.

Du fait des spécificités des dispositifs médicaux, ces essais contrôlés randomisés conventionnels peuvent être difficiles à mettre en place sur le plan pratique.

D'autres types de plans expérimentaux ont été décrits, en particulier dans la chirurgie, domaine dans lequel on se heurte à des difficultés similaires à celles rencontrées dans le champ de l'évaluation des dispositifs médicaux⁴.

7.2.1. Plans expérimentaux

7.2.1.1. « Zelen » ou « *randomised consent design trial* »

- **Contexte**

La recherche biomédicale et les essais randomisés classiques en particulier nécessitent d'informer le patient sur les bénéfices et les risques encourus dans l'essai. Habituellement, la randomisation n'est réalisée qu'après consentement éclairé du patient. Ainsi, les patients ne donnant pas leur consentement ne sont pas inclus. De plus, certains patients ayant initialement donné leur consentement peuvent se rétracter après avoir eu connaissance des résultats de la randomisation. En effet, le patient peut préférer recevoir un nouveau traitement prometteur plutôt que le traitement de référence. Afin de pallier ces problèmes d'inclusion, Zelen a proposé que le consentement du patient ne soit demandé que pour le nouveau traitement et non pour le traitement de référence (32).

- **Principe**

La randomisation « de Zelen » consiste à randomiser les patients sans avoir préalablement recueilli leur consentement éclairé. Seuls les patients randomisés dans le groupe nouveau traitement devront signer le consentement éclairé. En cas de refus, le patient recevra le traitement de référence (33).

Trois possibilités ont été proposées par Zelen.

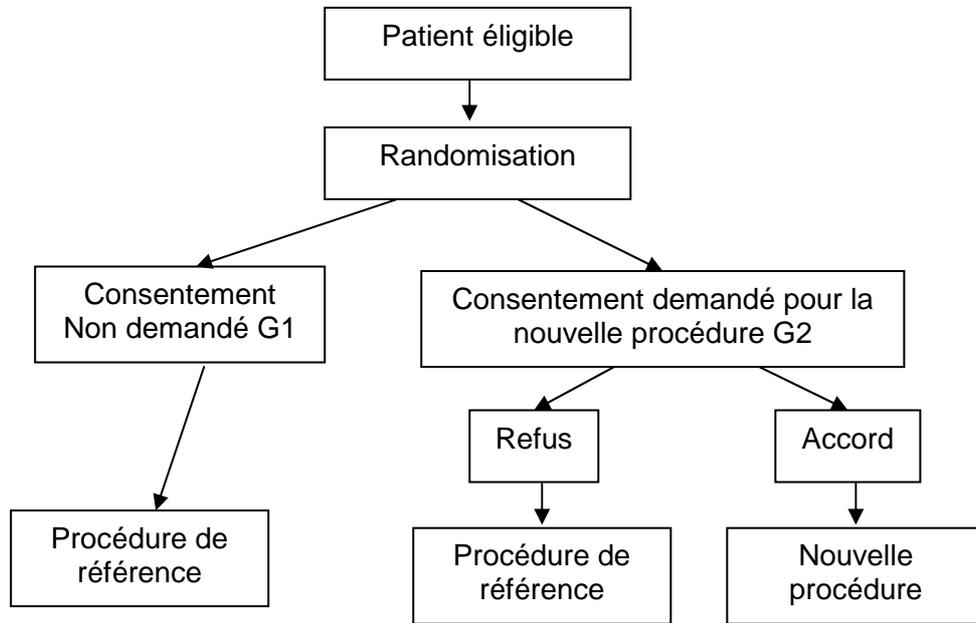
Dans la première (Figure 2), en cas d'éligibilité, le patient est randomisé dans l'un des 2 groupes suivants : le groupe G1 dans lequel le consentement n'est pas demandé, et le groupe G2 dans lequel le consentement est demandé (32). Les patients du groupe G1 reçoivent le traitement de référence et sont informés sur le traitement. Le consentement est demandé dans le groupe G2, pour recevoir le traitement expérimental. Si les patients du groupe G2 donnent leur accord, ils recevront le nouveau traitement. Dans le cas contraire, ils recevront le traitement de référence (32). L'analyse nécessitera alors que le groupe G1 (traitement de référence) soit comparé avec le groupe G2 (recevant soit le nouveau traitement, soit le traitement de référence). La comparaison doit être faite avec tous les patients du groupe G2, quel que soit le traitement reçu. Mais si seule une petite

³ « parallel groups » ou « parallel arms »

⁴ <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf>

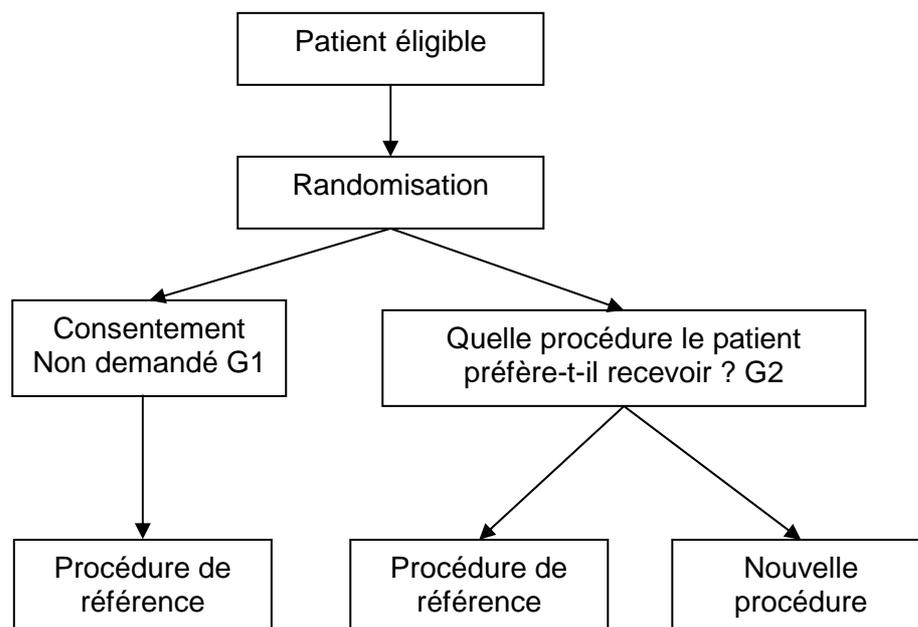
proportion des patients donne son accord pour recevoir le nouveau traitement, ce plan expérimental sera inutile pour évaluer l'efficacité du nouveau traitement. Cependant, une forte proportion de refus pour le nouveau traitement peut également indiquer qu'il est prématuré d'utiliser le traitement expérimental dans un essai clinique (32).

Figure 2 : Prérandomisation d'après Zelen, cas 1 (32)



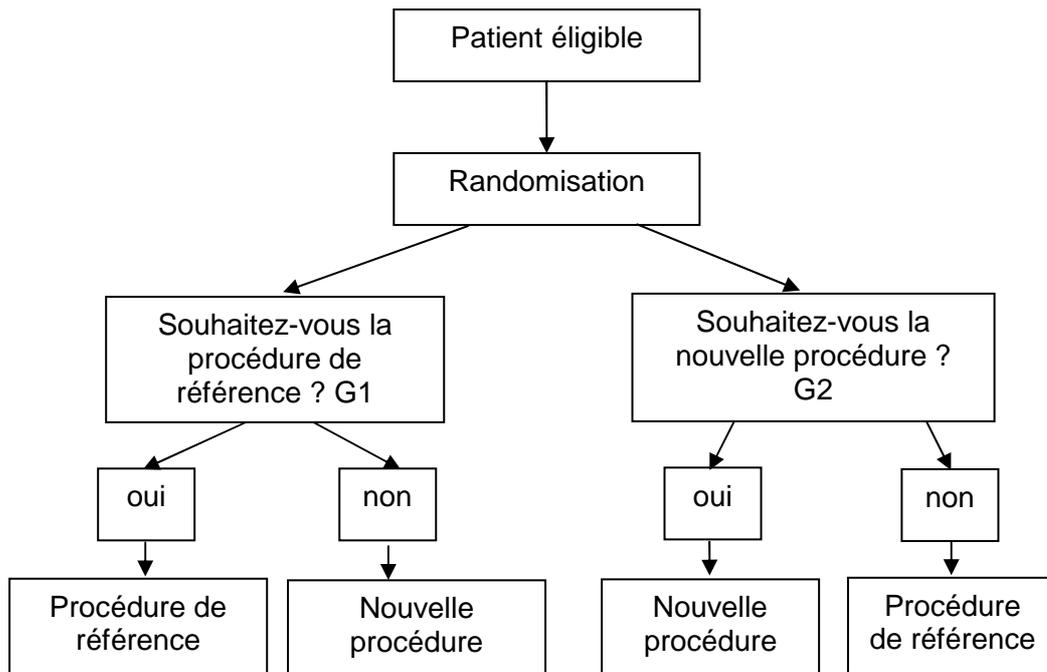
Dans la deuxième possibilité (Figure 3), lorsque le patient est randomisé dans le groupe G2, il est interrogé sur le traitement qu'il préfère recevoir. Cette approche est à privilégier en cas de traitement particulièrement incapacitant (ex : cancer de la prostate) (32).

Figure 3 : Prérandomisation d'après Zelen, cas 2 (32)



Enfin, la 3^{ème} possibilité proposée par Zelen est aussi appelée « double consentement » (Figure 4). Le patient est randomisé pour le traitement A ou le traitement B. S'il est randomisé dans le groupe traitement A, on lui demande s'il souhaite recevoir le traitement A. S'il refuse, il reçoit le traitement B ou un autre traitement. S'il est randomisé dans le groupe traitement B, on lui demande s'il souhaite recevoir le traitement B. S'il refuse, il reçoit le traitement A, ou un autre traitement (34). Cette alternative est particulièrement intéressante lorsqu'il n'y a pas de traitement de référence établi.

Figure 4 : Double consentement d'après Zelen, cas 3 (34)



La comparaison est effectuée entre les deux groupes, quel que soit le traitement finalement reçu (34). La comparaison porte sur le groupe G1 et sur le groupe G2 dans son ensemble, quel que soit le traitement reçu (32).

- **Avantages**

Le patient connaît le traitement qui lui sera administré avant de donner son consentement. Ceci peut être un avantage par rapport aux essais conventionnels (hors cas de double consentement), dans lesquels le patient peut se rétracter une fois le traitement attribué. De plus, ce type d'étude exige une décision du patient uniquement lorsque celui-ci doit recevoir le traitement expérimental (35). Ceci permet de limiter le stress dans certaines situations comparant une nouvelle technique au traitement standard. De plus, le recrutement des patients est facilité même en cas de forte préférence d'un des traitements de la part des patients. Ce type d'étude n'influe pas sur la confiance que le patient a en son médecin (32), dans la mesure où le médecin présente un seul traitement au patient. Aux yeux du patient, il n'apparaît pas comme celui qui attribue au hasard un traitement.

Au total, ce plan d'étude peut permettre d'inclure plus de patients que dans un essai conventionnel (32).

- **Inconvénients**

Un essai randomisé avec pré-randomisation soulève des problèmes éthiques spécifiques, en particulier pour le plan avec consentement simple, pour lequel on dispose d'un traitement de référence. Ceci pose la question de savoir si le patient doit savoir comment son traitement a été choisi (36). Du fait de l'absence de consentement éclairé dans le groupe G1, ce plan ne peut pas être utilisé dans les essais contre placebo (37). De plus, de nombreux essais cliniques nécessitent des visites de suivi fréquentes pour le recueil des données. Ces visites sont généralement plus fréquentes que dans les modalités de suivi habituel. Ceci ne pose pas de problèmes pour les patients du groupe G2 ayant fourni leur consentement. En revanche, ceci pose des problèmes pour les patients du groupe G1 lorsqu'ils seront informés de visites plus fréquentes ou plus invasives pour la collecte des données et des surveillances (37)

Comme la comparaison se fait entre l'ensemble des patients randomisés dans le groupe G1 et l'ensemble des patients randomisés dans le groupe G2, l'efficacité du nouveau traitement peut être masquée si beaucoup des patients de G2 choisissent le traitement de référence (37). Ainsi, le fait de comparer le groupe G1 au groupe G2 quel que soit le traitement reçu dilue l'effet mesurable du nouveau traitement (32). Ainsi, il existe un risque de ne pas pouvoir conclure.

Enfin, le double aveugle est impossible (32), et le biais de sélection constitue l'inconvénient majeur de cette stratégie. En effet, les biais de sélection sont accentués par une sous-représentation des patients à mauvais pronostic dans le groupe expérimental, car il risque de comporter un taux de refus de consentement proportionnel à la sévérité de l'atteinte⁵.

Au total, ce type de plan s'apparente plus aux études observationnelles.

- **Contraintes**

Les patients du bras contrôle doivent être informés que des données sont recueillies. Il est également important de comparer les caractéristiques des patients du groupe 1 avec celles des patients du groupe 2 recevant le traitement de référence suite à leur refus de recevoir le nouveau traitement, afin d'identifier des biais de sélection potentiels (32).

- **Champs d'application**

Ce plan expérimental est particulièrement intéressant lorsque les patients recevant le traitement de référence ne nécessitent pas de visites supplémentaires et que le décès est le seul critère de jugement (37).

Il peut être utilisable dans les domaines de la chirurgie, des études interventionnelles, ou des dispositifs médicaux (38).

Du fait de l'absence d'information donnée dans le groupe randomisé pour recevoir le traitement conventionnel, ce type d'étude est débattu sur le plan éthique. Très peu d'études l'utilisent en France. En pratique, il est surtout utilisé pour des pathologies dans lesquelles la décision du traitement est psychologiquement difficile pour le patient (cancérologie, soins palliatifs, choix du traitement pour son enfant...).

⁵ Or il n'y aura pas de sous-représentation dans le groupe témoin puisque le consentement ne sera pas demandé dans ce groupe.

7.2.1.2. Essais basés sur l'expertise⁶

- **Contexte**

Les essais contrôlés randomisés classiques peuvent être difficiles à mettre en place en chirurgie du fait de problèmes d'acceptabilité pour le chirurgien. Une mauvaise acceptabilité peut gêner le recrutement de patients dans un essai clinique conventionnel (9). De plus, un chirurgien pourrait inconsciemment réaliser de manière plus méticuleuse la procédure chez les patients randomisés pour l'intervention qu'il préfère, ou suivre les patients d'une manière différente que les chirurgiens moins expérimentés (9,39). Le protocole de l'étude doit par conséquent être précis et suivi rigoureusement.

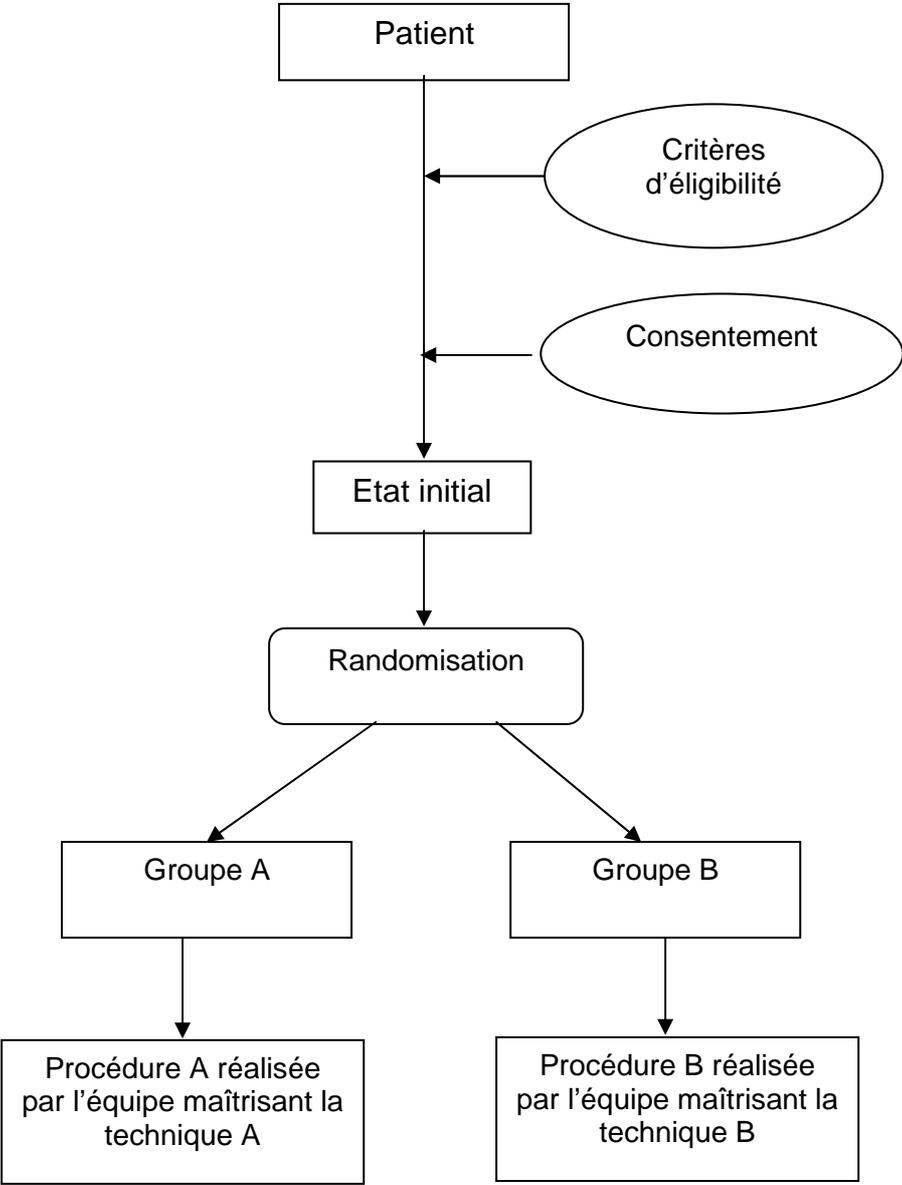
Ces éléments soulignent les limites de l'essai randomisé conventionnel classique dans le domaine interventionnel. Pour tenter de pallier ces limites, les essais contrôlés randomisés basés sur l'expertise (« expertise-based randomized controlled trials ») ont été développés.

- **Principe**

Contrairement à l'essai clinique conventionnel, dans lequel le patient est randomisé pour recevoir soit l'intervention A, soit l'intervention B, réalisées par une même équipe, l'essai contrôlé randomisé basé sur l'expertise consiste à randomiser les patients auprès du chirurgien ou de l'équipe qui maîtrise l'intervention dédiée (Figure 5) (9). Le chirurgien ou l'équipe ne réalise qu'une seule procédure, dont ils ont la maîtrise et pour lesquels ils sont experts. Il faut s'assurer que la maîtrise de la procédure est comparable dans chaque groupe (27). Ce type d'essai peut limiter l'acceptabilité pour le patient qui doit accorder sa confiance à deux chirurgiens ou à deux équipes.

⁶ ou « *expertise-based randomized controlled trials* »

Figure 5 : Schéma représentant la randomisation dans les essais contrôlés randomisés basés sur l'expertise, d'après (27)



- **Avantages**

- *Amélioration de la validité interne*

Les essais contrôlés randomisés fondés sur l'expertise offrent une meilleure validité interne, (39) pour les raisons suivantes :

- *Absence de biais d'expertise différentielle*

Les essais contrôlés randomisés basés sur l'expertise permettent de réduire le biais d'expertise différentiel qui existe dans les essais conventionnels (40).

- *Limitation des biais liés à l'absence d'aveugle*

Ce type d'essai permettrait également de limiter les biais de suivi différentiels liés à l'absence d'aveugle. En effet, chaque chirurgien ne réalisant que l'intervention dans laquelle il est spécialisé, le risque de différences entre les patients concernant l'intervention, et les facteurs annexes seraient plus limités que dans un essai conventionnel dans lequel les chirurgiens pourraient suivre de manière différente les patients entre les différents groupes (39). Cependant rien ne garantit que les patients soient suivis et pris en charge de la même façon dans les deux groupes, d'où la nécessité de spécifier et de standardiser ces modalités dans le protocole avant la mise en place de l'essai.

- *Réduction des écarts au protocole*

Afin de pouvoir évaluer au mieux l'efficacité d'un nouveau traitement, les patients doivent recevoir le traitement pour lequel ils ont été randomisés. Si ce n'est pas le cas, ces écarts au protocole altèrent la validité interne de l'essai clinique. Devereaux et al (39) ont montré dans un essai clinique conventionnel sur le traitement des fractures tibiales (alésage vs non alésage) des écarts au protocole différentiel entre les deux bras, avec une plus grande proportion de patients randomisés dans le groupe sans alésage (techniquement plus délicate). Ces écarts au protocole sont d'autant plus fréquents que les chirurgiens pratiquent peu l'intervention dans laquelle le patient a été randomisé que dans le cas contraire (41-43). De ce fait, le risque d'écart au protocole différentiel est réduit dans les essais basés sur l'expertise, le chirurgien réalisant uniquement l'intervention qu'il a l'habitude de pratiquer (39).

- *Amélioration de la faisabilité*

La faisabilité d'un essai randomisé basé sur l'expertise est meilleure que celle des essais conventionnels. En effet, les chirurgiens n'ont pas besoin d'avoir été formés à deux procédures et seront plus facilement recrutés avant de débiter l'essai clinique (39).

- *Meilleure acceptabilité par les chirurgiens*

Une enquête transversale a montré que 58% des chirurgiens orthopédiques préféraient participer à des essais contrôlés basés sur l'expertise, contre 17% pour les essais cliniques contrôlés randomisés conventionnels. La préférence était influencée par l'expertise des chirurgiens (9). De plus, l'acceptabilité vis-à-vis de problèmes éthiques est améliorée dans la mesure où le chirurgien ne réalise que l'intervention qu'il a l'habitude de pratiquer, qu'il préfère, et pour laquelle il est « expert » (27,39). Au total, les chirurgiens sont moins réticents à participer à un essai clinique basé sur l'expertise plutôt que sur un essai conventionnel (27).

- ***Inconvénients***

Comme dans les essais conventionnels portant dans ce domaine, l'aveugle n'est pas possible (39), et le risque de biais de mesure est majeur.

Plus spécifiquement, il peut parfois être difficile de savoir si la supériorité d'une technique par rapport à une autre est réelle, ou si elle est liée à l'expertise du chirurgien qui la réalise (9). Il en découle la nécessité de recruter des chirurgiens extrêmement compétents dans chaque groupe. Ceci peut entraîner à son tour des délais d'attente et un temps d'inclusion plus long comparativement à un essai conventionnel (6,40).

- ***Contraintes :***

Ce type de schéma nécessite au moins la présence d'un expert de chacune des interventions randomisées dans chaque centre (9). De plus, la consultation initiale de pré-inclusion doit être réalisée par une personne neutre qui déterminera l'éligibilité des patients (9).

Enfin, pour qu'un chirurgien puisse participer à l'essai, il doit avoir atteint un niveau de compétence suffisant (essai explicatif), ou une phase de plateau dans la courbe d'apprentissage (essai pragmatique), afin d'éviter un biais différentiel lié à l'expertise du chirurgien (39). Ceci suppose de ne pas évaluer une nouvelle intervention trop précocement, au risque de refléter un défaut d'apprentissage de la technique plutôt que l'efficacité réelle de l'intervention.

- ***Champs d'application***

Les essais basés sur l'expertise sont potentiellement intéressants dans toutes les situations où il y a un intérêt avéré de tenir compte de l'existence d'une courbe d'apprentissage. En effet, lorsque l'utilisation d'un nouveau dispositif médical nécessite d'acquérir de l'expérience, une évaluation trop précoce peut davantage refléter le manque d'expérience de l'opérateur que l'effet réel du nouveau dispositif. Ce type d'essai pourrait être utilisé :

- lorsque les deux techniques sont très différentes et nécessitent chacune une courbe d'apprentissage ;
- lorsque la maîtrise de la technique a une influence importante sur le résultat.

7.2.1.3. « *Tracker trial design* » ou essais suiveurs

- **Contexte**

Les dispositifs médicaux sont fréquemment commercialisés avant la réalisation d'études approfondies sur leur efficacité. Ils sont de plus sujets à des évolutions technologiques importantes (44). L'essai « *tracker* » ou suiveur a été développé afin de pouvoir prendre en compte ces évolutions technologiques ou l'amélioration de la procédure au cours de l'essai (6).

- **Principe**

Des modifications au cours de l'essai sont autorisées, enregistrées et prises en compte dans l'analyse statistique. Les variations dans le schéma d'étude sont également permises (6).

- **Avantages**

L'essai suiveur ou « *tracker* » permet de s'adapter aux évolutions de la technologie et de la pratique clinique. Il permet une évaluation précoce, respectant ainsi le principe d'équilibre⁷ (absence d'arguments pour conseiller un traitement plutôt qu'un autre), l'utilisation maximale des différentes données disponibles, et permettent d'évaluer rapidement des techniques nouvelles ou innovantes ou potentiellement dangereuses (44).

- **Inconvénients**

L'analyse est complexe du fait de la prise en compte de l'expérience de l'opérateur et l'évolution du traitement. Ce type d'essai fait appel à des méthodes plus sophistiquées que les essais conventionnels. Il nécessite des budgets flexibles pour prendre en compte notamment la durée de l'étude (44).

L'organisation pratique est difficile (6,44), d'autant plus quand la technologie évolue, car la plupart des opérateurs sont alors peu expérimentés (44).

Il existe peu de données dans la littérature sur les analyses spécifiques à réaliser dans ce type d'essai.

- **Contraintes**

Les essais « *tracker* » doivent être flexibles et inclure les nouveaux traitements au fur et à mesure de leur apparition. La révision régulière du protocole est nécessaire pour pouvoir intégrer de nouveaux bras de traitements, nouveaux ou émergents, ou au contraire pour supprimer certains bras. Tous les opérateurs et les centres doivent être inclus indépendamment de leur niveau d'expérience. Cependant, il est nécessaire de prendre en compte cet élément dans l'analyse (44).

- **Champ d'application**

Actuellement, il existe peu d'exemples disponibles dans la littérature sur ce type d'étude. Il semblerait présenter plutôt un intérêt en phase préliminaire, pour écarter des technologies dangereuses, et donc en amont.

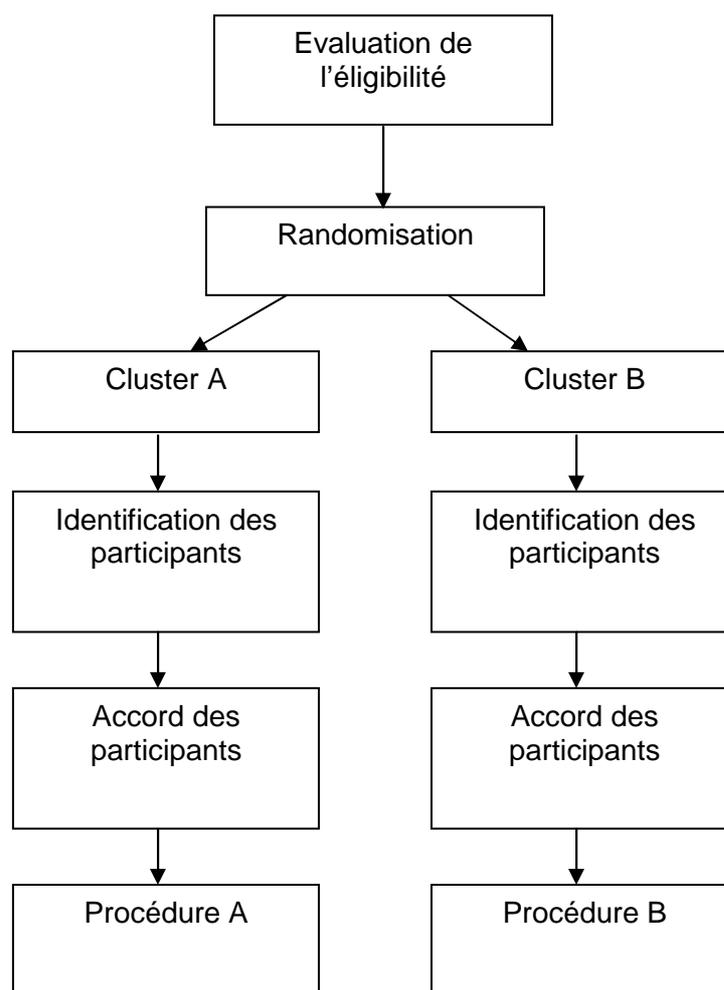
⁷ également appelé « équipose »

7.2.1.4. Essais en « cluster »

- **Principe :**

Les essais en cluster consistent à randomiser des groupes (clusters) d'individus (par centre, hôpital, service), plutôt que les individus directement (Figure 6). En pratique, on tire au sort la technique que doit utiliser un centre. Par exemple, l'hôpital A utilisera systématiquement la technique 1, et l'hôpital B utilisera systématiquement la technique 2.

Figure 6 : Schéma représentant les essais en cluster d'après (45)



Deux types de plan d'essai en cluster ont été décrits: les essais en cluster complet, ou les essais en cluster avec recrutement actif (46). Dans les essais en cluster complet, une personne dans chaque centre est en charge du cluster et se voit attribuer la procédure à tester.

Avantages

La randomisation de clusters plutôt que d'individus permet de prévenir une éventuelle contamination entre deux techniques disponibles dans un même centre (46). Ainsi, lors de l'utilisation d'un antiseptique en post-opératoire, le risque d'administrer l'antiseptique A au lieu de l'antiseptique B est beaucoup plus limité lorsque seul l'un des deux est utilisé dans l'hôpital, plutôt que si les deux sont disponibles). Elle est également plus simple sur le plan logistique. Ce plan expérimental est particulièrement utile pour la comparaison de stratégies globales ou multimodales de prise en charge.

▪ **Inconvénients**

Ce type d'essai peut manquer de puissance lorsque l'effet cluster n'est pas pris en compte dans le calcul du nombre de sujets nécessaires (46).

De plus, du fait de leurs modalités de recrutement et de suivi, les essais en cluster peuvent compromettre la comparabilité des groupes. En effet, certains participants peuvent être inclus dans un cluster alors qu'ils auraient dû être affectés dans l'autre. Or, en l'absence de connaissance sur le cluster dans lequel chaque patient aurait dû être inclus, l'analyse en intention de traiter est difficile et met en péril la comparabilité des groupes qui devrait être obtenue tant au niveau individuel qu'au niveau cluster dans ce type d'essai (46).

De plus, un recrutement différentiel peut exister entre les clusters pour les essais en cluster avec recrutement actif, en termes de nombre de patients et de caractéristiques des patients inclus (46). Un facteur important de validité est l'exhaustivité des patients analysés. Enfin, il existe un risque de contamination entre clusters en cas de transfert de patients entre services.

Pour les essais en cluster complet, il peut y avoir des clusters vides lorsque la personne en charge du cluster refuse de participer juste après avoir eu connaissance du résultat de la randomisation (46). Aucune méthode statistique ne permet de limiter ce biais. De plus, l'influence de l'unité de randomisation (individu ou groupes d'individus) est beaucoup plus forte dans un essai en cluster que dans un essai conventionnel (46).

Pour les essais en cluster avec recrutement actif, il peut également y avoir des responsables du cluster inactif, et au pire de manière différentielle, car ils n'adhèrent pas au résultat de la randomisation. De plus, le recrutement des patients s'effectuant après la randomisation, les participants sont sélectionnés par une personne qui a connaissance du groupe dans lequel ils seront randomisés. Les patients ne donnent pas leur consentement pour la randomisation, mais pour leur participation dans un groupe pré-déterminé, ce qui peut induire un biais de sélection et un recrutement différentiel (46). Le recrutement différentiel pourrait être assimilé au biais de non réponse, mais le nombre de participants éligibles est inconnu, ce qui rend cette estimation complexe (46).

Les essais en cluster soulèvent également des questions éthiques, puisqu'ils supposent 2 niveaux de consentement : celui des professionnels de santé, puis l'information des patients sur le fait qu'ils participent à une étude (47). Ainsi, les patients n'auront pas le choix du traitement. Dans les essais où un nouveau traitement est comparé au traitement de référence, il n'est pas nécessaire d'informer les patients de ce dernier groupe, dans la mesure où leur traitement est celui habituellement utilisé. Cependant, en cas de mesures supplémentaires pour évaluer l'efficacité, les patients doivent avoir été préalablement informés et avoir donné leur accord. Ceci entraîne un problème supplémentaire pour l'aveugle du patient (47).

- **Contraintes**

Le calcul du nombre de sujets nécessaires doit prendre en compte l'effet cluster, ce qui conduit à une augmentation du nombre de sujets nécessaires (46). L'analyse statistique doit également tenir compte de cet effet cluster (46).

En effet, il existe dans ce type d'étude une structure hiérarchique des données : un échelon cluster, et un échelon individuel. Il est nécessaire de tenir compte du fait que les individus au sein du même cluster vont tendre à plus se ressembler que les individus entre différents clusters.

Pour les essais en cluster complet, il faut s'assurer que la personne en charge des clusters adhère au protocole de l'étude avant la randomisation (46).

Afin de limiter les biais de sélection et le recrutement différentiel dans les essais en cluster avec recrutement actif, Puffer et al (48) ont proposé l'identification et l'inclusion complète des participants avant la randomisation d'un cluster. Cette stratégie, qui prévient à la fois les clusters vides et le biais de sélection, devrait être impérative, même si elle est très difficile à assurer. En effet, elle ne peut pas être systématiquement mise en oeuvre pour des raisons logistiques. D'autres alternatives ont été proposées, comme le recrutement en aveugle du groupe d'allocation ou la randomisation des clusters uniquement lorsque le premier participant a été inclus (« *index case* »). Cette dernière option peut permettre d'éviter les clusters vides, mais n'évite pas le recrutement différentiel (46).

- **Champs d'application**

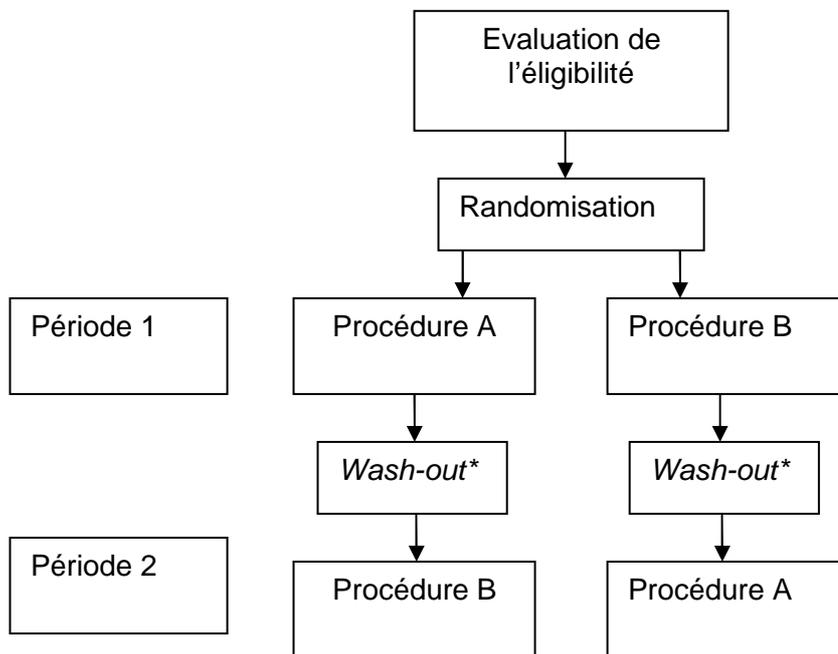
Les essais en cluster sont essentiellement utilisés pour les interventions organisationnelles, comportementales ou encore les programmes de promotion pour la santé. Dans la mesure où ils sont susceptibles d'améliorer la faisabilité ou les aspects pratiques, ils pourraient être envisagés dans le champ des dispositifs médicaux ou des techniques interventionnelles ou chirurgicales. Ainsi, ils ont déjà été utilisés dans le domaine de l'évaluation de différentes stratégies en hygiène hospitalière.

7.2.1.5. Essais en « cross over »

- **Principe**

Par opposition aux essais en groupes parallèles, dans lesquels chaque patient ne reçoit qu'un seul traitement, les essais en cross-over consistent à randomiser la séquence dans laquelle chaque traitement est administré (Figure 7). Chaque patient reçoit ainsi les deux traitements (49). Ainsi, les différents traitements sont comparés sur le même patient au cours de différentes périodes (50).

Figure 7 : Principe du *cross-over*



* *wash-out* = fenêtre thérapeutique ou phase de sevrage

- **Avantages**

Le sujet étant son propre témoin, à précision égale et pour des mesures corrélées entre les périodes, le nombre de sujets nécessaires est réduit (50), ce qui est particulièrement intéressant dans le cas de maladies rares. De plus, le patient peut exprimer sa préférence pour l'un ou l'autre des traitements (49). Enfin, les deux groupes sont parfaitement comparables puisque chaque sujet reçoit les deux traitements

- **Inconvénients**

La préoccupation majeure des essais en cross-over est le risque d'effet dit « *carry-over* ». Cet effet correspond au fait que le traitement administré en 1^{er} a un effet rémanent au cours de la deuxième période. Un autre inconvénient est que les essais en cross-over sont inadaptés lorsque le critère de jugement étudié est la guérison ou le décès (49), ou lorsqu'il y

a une courbe d'apprentissage. Dans le cas où les mesures ne seraient pas corrélées entre les périodes, la puissance d'un essai en cross-over est deux fois plus faible qu'un essai en bras parallèles à effectif identique.

- **Contraintes**

Les essais en cross-over supposent que la pathologie étudiée soit stable, qu'il n'y ait pas d'effet rémanent du premier traitement reçu lorsque le second traitement est administré et que le critère de jugement puisse se répéter. La réalisation d'une période de « *wash-out* » (fenêtre thérapeutique ou phase de sevrage) entre les deux périodes d'administration du traitement permet de réduire ce risque (49).

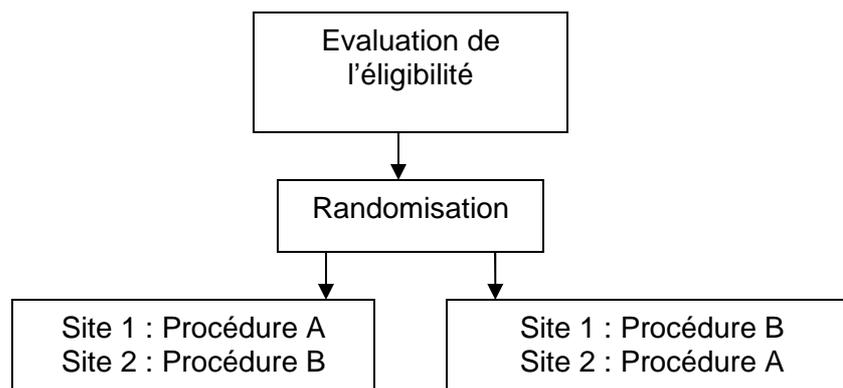
Il est nécessaire de rechercher une interaction entre le traitement et la période (49) ou effet carry-over). Si celui-ci est présent, seules les données issues de la première période seront analysées.

Des tests spécifiques pour séries appariées doivent être utilisés pour tenir compte du fait que chaque sujet est son propre témoin. Enfin, le traitement ne doit pas avoir d'effet irréversible.

- **Cas particuliers de « cross-over »**

Deux traitements peuvent être administrés simultanément chez un même patient. Par exemple deux traitements topiques, appliqués simultanément sur deux sites cutanés tirés au sort (Figure 8). Ce type de protocole suppose l'absence d'interaction possible entre les deux sites d'application, comme la diffusion systémique du produit administré.

Figure 8 : Cas particulier du cross-over lorsque deux traitements sont appliqués simultanément



- **Champs d'application**

Le cross-over est utilisé dans l'évaluation de traitements dans le cadre de maladies chroniques, en dermatologie et en cosmétologie. Il pourrait avoir une utilisation plus répandue pour l'évaluation de certains dispositifs médicaux.

7.2.2. Analyses intermédiaires

Les analyses intermédiaires ont pour objectif de pouvoir arrêter précocement l'essai si un des traitements évalués s'avère plus efficace, ou en cas de problème de tolérance dans l'un des bras. Afin de ne pas engendrer d'augmentation du risque de conclure à tort à une différence qui n'existe pas (inflation du risque alpha), l'utilisation de méthodes adaptées est nécessaire.

Différentes stratégies d'analyse permettent de réaliser des analyses intermédiaires (devant avoir été planifiées avant le début de l'étude).

Parmi elles, on peut citer les essais séquentiels, et les essais adaptatifs.

Les méthodes bayésiennes, développées plus loin, constituent également une stratégie d'analyse des données.

7.2.2.1. Essais séquentiels

- **Contexte**

Les essais séquentiels font partie des techniques d'analyse intermédiaire.

Les essais séquentiels ont été développés afin de permettre de ne pas donner un traitement moins efficace à des patients plus longtemps que nécessaire (51).

- **Principe**

Les différentes étapes sont planifiées indépendamment des données collectées auparavant (52).

Différentes méthodes d'analyses séquentielles ont été décrites. La méthode séquentielle classique repose sur l'analyse de paires de patients (53), les analyses séquentielles en groupe sur l'analyse tous les n sujets (54-58), les méthodes flexibles (59,60) permettent de choisir à quel moment faire les analyses (57,58), et les analyses séquentielles continues comme le test séquentiel de rapport de probabilité et le test triangulaire (57,58,61).

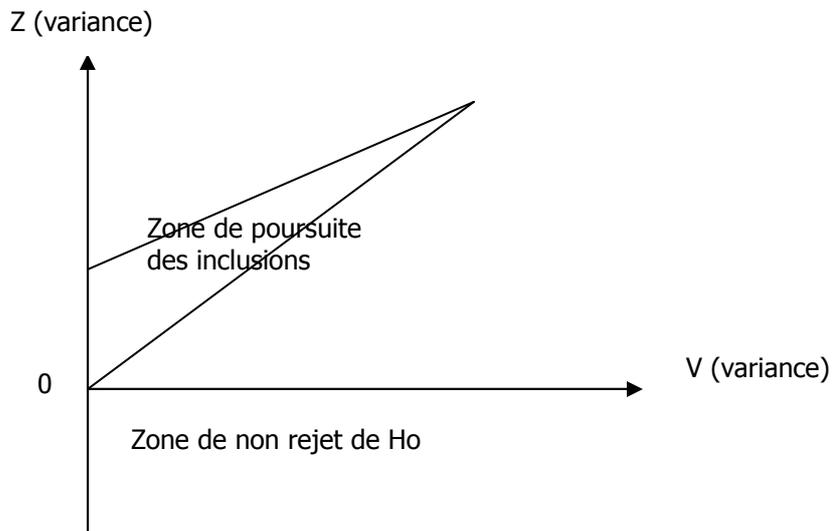
Les essais séquentiels sont des essais contrôlés randomisés dans lesquels les résultats de l'essai concernant les patients inclus sont examinés avant la randomisation de nouveaux patients, et une décision est prise concernant la poursuite ou non de l'étude. Le but est d'éviter les difficultés éthiques sur la poursuite de l'étude si la différence entre traitements s'avère plus importante que celle qui était attendue (62).

Ces analyses reposent sur des méthodes graphiques. Les frontières sont déterminées avant l'essai, sur la base de l'hypothèse alternative, du risque alpha et du risque beta. L'inclusion des patients se poursuit tant que la droite reste comprise entre les deux frontières. Si la frontière supérieure est franchie, l'hypothèse nulle est rejetée, tandis que si la frontière inférieure est franchie, l'hypothèse nulle n'est pas rejetée (51). Dans ces deux cas, les inclusions sont stoppées.

La figure 9 et la figure 10 représentent les règles de décision graphique pour arrêter ou pour poursuivre les inclusions.

Les spécificités de ces différentes méthodes sont présentées dans le tableau 4.

Figure 9: Règles de décision graphique pour le test triangulaire (plan fermé) d'après (63,64).
Analyse séquentielle des essais thérapeutiques



Le nombre maximal d'analyses est limité car la zone de poursuite des inclusions est fermée.

Figure 10 : Règle de décision graphique pour le test séquentiel de rapport des probabilités (plan ouvert)

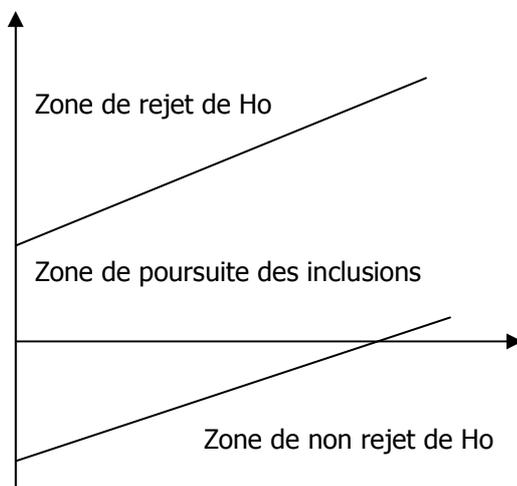


Tableau 4 : Principe, avantages et inconvénients spécifiques des différentes méthodes d'analyse séquentielle

Nom	Principe	Avantages	Inconvénients
Méthode séquentielle classique			
Armitage 1975 (53)	<ul style="list-style-type: none"> - Appariement : soit le patient est pris pour son propre témoin, soit il s'agit de 2 sujets distincts avec prise en compte de facteurs pronostiques ou tirés au sort - L'analyse se fait tous les 2 patients pour les critères de jugement obtenus rapidement après une inclusion 		<ul style="list-style-type: none"> - application lourde - nécessité analyses fréquentes - nécessité d'avoir un critère de jugement unique et obtenu rapidement - peu adapté aux données censurées - test ouvert : on peut ne jamais conclure
Méthodes séquentielles de groupe : analyse tous les n sujets			
Pocock 1977, Pocock 1982 (54,65)	<ul style="list-style-type: none"> -Même risque alpha pour toutes les analyses, fixé en fonction du nombre d'analyses intermédiaires -Analyses intermédiaires réalisées après inclusion d'un nombre de patients fixe 	<ul style="list-style-type: none"> -Simple -Permet de conclure plus vite que O'Brien et Fleming (57,58,65) -Permet de conclure plus vite que Peto (57,58,65) 	<ul style="list-style-type: none"> -Ne permet pas de conclure précocement en cas d'équivalence entre 2 traitements -Nombre d'analyses intermédiaire fixé dès le départ -Parfois résultats discordants selon la réalisation ou non d'une analyse intermédiaire (57,58)-Nécessite le nombre de sujets nécessaires le plus important (66)
O'Brien et Fleming 1979 (55)	Augmentation du risque alpha au fur et à mesure des analyses intermédiaires (57)	Permet de conclure plus vite que Peto	Nombre d'analyses fixé dès le départ
Peto (56)	Seuil $p < 0,001$ pour chaque analyse	Plus souple en termes de nombre d'analyses intermédiaires (57)	
Méthode des successions Falissard 1992 (58)	On impose que les résultats soient significatifs à 0.005 pour une succession d'analyses intermédiaires.		

Méthodes flexibles			
Lan der Mets (67)	Alpha différent à chaque analyse, déterminé en fonction des analyses intermédiaires passées et en cours	Nombre d'analyses intermédiaires non prédéfini	-Risque d'augmentation du NSN -Faible probabilité de conclure en cours d'essai (51)
Méthode semi-séquentielle (57)	adaptation de la méthode de succession avec liberté de la date et du nombre d'analyses intermédiaires (Falissard, Lelouch, Resp)		
Analyse séquentielle continue (59), SPRT et TT (61,63)			
Test triangulaire		Méthode de choix Test fermé Conclusion plus précoce	
Test séquentiel de rapport de probabilité			Test ouvert : risque de ne pas conclure et d'une durée longue de l'essai

- **Avantages liés aux méthodes séquentielles**

Ce type d'analyse intermédiaire permet de répéter les analyses sans inflation du risque alpha (éviter de conclure à tort à une différence qui n'existe pas) et tout en conservant la puissance. D'un point de vue éthique ces méthodes ont l'intérêt de permettre un arrêt de l'essai en fonction des résultats des inclusions précédentes, et d'autoriser une conclusion précoce de l'étude (62).

En outre, le nombre moyen de sujets nécessaires dans ce type d'essais est plus faible que dans les essais conventionnels, indépendamment de la taille de l'effet ou de la puissance (51,68).

Benichou et al ont plus spécifiquement comparé le test triangulaire, le test séquentiel de rapport de probabilités et l'analyse séquentielle groupée. Ils ont montré que ces trois méthodes offraient des propriétés statistiques correctes, mais que le test triangulaire permettait une conclusion plus précoce quelle que soit la taille de la différence entre les deux groupes comparés (63,64,69). Le test triangulaire est donc préférable à l'analyse séquentielle groupée qui ne permet pas de conclure précocement en cas d'équivalence des deux traitements comparés.

- **Inconvénients des méthodes séquentielles**

Les méthodes séquentielles ont des contraintes pratiques. D'une part, elles permettent de comparer seulement deux traitements. D'autre part, la règle d'arrêt des inclusions repose sur un seul critère. Il est donc nécessaire de trouver un critère de jugement unique sur lequel le traitement sera évalué (62). Un arrêt précoce de l'essai risque d'engendrer un manque de puissance pour les critères de jugement secondaires, ou d'influer sur la mesure de la fréquence des effets indésirables, ce qui est extrêmement dommageable. De ce fait, si on s'intéresse à la fréquence des effets indésirables, il est nécessaire d'avoir une puissance suffisante (51). Enfin, le nombre de sujets au départ n'est pas connu, et les intervalles de confiance sont plus larges que ceux obtenus avec un essai conventionnel (51).

La validité externe des essais séquentiels a également été mise en question du fait de résultats pouvant être obtenus sur un faible échantillon, potentiellement avec un biais de sélection, mais cet argument a été réfuté par Van der Lee et al du fait de la randomisation (51).

L'approche par frontières suppose que le temps entre l'inclusion et la mesure du critère de jugement soit rapide. Si beaucoup de patients sont inclus sur une courte période, l'utilisation des données obtenues chez seulement une partie d'entre eux pour décider de l'arrêt des inclusions n'est pas satisfaisante (51). Cela ne conduit pas à une réduction du nombre de sujets nécessaires, mais de la durée de l'essai. Les frontières sont définies d'après les connaissances a priori et ne peuvent être changées selon les données collectées avec les analyses séquentielles (51).

- **Contraintes**

La confidentialité des résultats à chaque étape s'impose. De ce fait, il est nécessaire de constituer un comité de surveillance des données indépendant, qui décidera des adaptations du nombre de sujets nécessaires et de la poursuite ou de l'arrêt de l'inclusion sans donner d'informations détaillées sur les résultats de l'analyse aux personnes impliquées dans l'essai (investigateurs, évaluateurs, patients) (51). L'inclusion doit être régulière pour pouvoir respecter la fréquence des analyses, avec un suivi de bonne qualité pour éviter des délais

dans la mise à jour des données. Des procédures permettant une surveillance régulière des patients et un retour rapide de l'information sont donc nécessaires. Il faut également avoir un suivi à long terme après l'arrêt de l'essai pour d'éventuels effets indésirables.

L'utilisation de méthodes spécifiques pour le calcul des intervalles de confiance (57,58) et la correction du biais des estimations sont également requises.

- ***Champs d'application***

Actuellement, ces techniques d'analyse ont un intérêt particulier dans le champ des maladies rares ou orphelines et en pédiatrie. De manière plus large, elles peuvent permettre d'arrêter plus précocement une stratégie thérapeutique qui s'avère inefficace, impliquant une perte de chance pour les patients qui y sont exposés.

7.2.2.2. Essais avec randomisation adaptative

- ***Contexte***

L'objectif de ces essais adaptatifs est de maximiser le nombre total de patients traités avec le meilleur traitement. Ainsi, au début de l'essai, les patients sont randomisés avec un ratio 1:1 dans chaque bras, puis, à mesure que les résultats commencent à montrer une différence entre les traitements, le ratio est modifié en faveur du groupe qui reçoit le traitement semblant être plus efficace (62).

L'essai adaptatif ne doit pas être confondu avec un essai dans lequel le ratio d'allocation est adapté selon les résultats préliminaires de l'essai (70).

- ***Principe***

Comme pour les essais séquentiels, les essais adaptatifs reposent sur la réalisation d'analyses intermédiaires. Mais dans ce type d'essai, le plan expérimental peut être changé sur la base des résultats observés lors des analyses intermédiaires (51) sans augmentation du risque alpha (71). Toute l'information collectée lors des différentes étapes est utilisée, et pas seulement celle de l'étape en cours (52). L'allocation du traitement d'un patient dépend des résultats obtenus avec les patients précédents (71).

En cas de modifications, une nouvelle phase de l'essai débute, et l'analyse porte sur les données accumulées dans cette nouvelle phase, et non plus sur l'ensemble des données. Les différentes adaptations possibles sont la réévaluation du nombre de sujets nécessaires, l'adaptation des critères de jugement, l'ajout ou la suppression d'analyses intermédiaires (51).

Les essais adaptatifs peuvent être en 2 phases ou en plusieurs phases (72). Les essais adaptatifs en 2 phases consistent dans un premier temps à sélectionner un échantillon de taille N1. Selon les résultats de cette première phase, soit l'étude s'interrompt, soit l'étude entre dans la seconde phase. Les essais adaptatifs en plusieurs phases sont une extension des essais adaptatifs en 2 phases.

- **Avantages**

Les essais adaptatifs permettent une grande flexibilité et la réduction du nombre de sujets nécessaires. De plus, toute l'information collectée est utilisée pour l'analyse (51,52). D'autre part, il a été montré que ce schéma d'étude permettait de traiter correctement un plus grand nombre de patients, en comparaison avec les essais séquentiels et les essais randomisés (62,73).

- **Inconvénients**

Un essai adaptatif requiert des contraintes logistiques plus grandes qu'un autre type d'essai. (52). Il exige également une grande vigilance concernant la ré-estimation du nombre de sujets nécessaires.

Des difficultés pratiques peuvent être rencontrées, mais évitées en convenant d'un essai en plusieurs phases, chaque phase utilisant un ratio différent pour la randomisation entre les groupes (62). Comme pour les essais séquentiels, il est nécessaire de définir un critère de jugement simple et univoque de l'efficacité (62). Van der Lee a posé le problème de la définition des règles d'arrêt ou de l'interprétation des résultats lorsque le critère de jugement principal a été changé au cours de l'étude (51).

La validité interne des essais adaptatifs est également remise en question. En effet, pour Armittage, ce type d'étude ne serait pas strictement randomisé et entraînerait les mêmes inconvénients que les témoins historiques (53,62).

Du fait de ces inconvénients, les essais adaptatifs ont été jugés peu intéressants par rapport aux essais séquentiels (62,74,75).

- **Contraintes de l'adaptatif**

Comme pour les autres méthodes faisant appel à des analyses intermédiaires, les résultats des essais adaptatifs peuvent être faussés par la connaissance des données intermédiaires de l'essai par les investigateurs. De plus, ces biais ne peuvent être corrigés par l'ajustement statistique, ce qui met en péril l'interprétation des résultats.

De ce fait, il est nécessaire de constituer un comité de surveillance des données indépendant, qui décidera des adaptations du nombre de sujets nécessaires et de la poursuite ou de l'arrêt de l'inclusion sans donner d'informations détaillées sur les résultats de l'analyse aux personnes impliquées dans l'essai (investigateurs, évaluateurs, patients) (51,76).

Enfin, l'enregistrement prospectif des hypothèses et des adaptations réalisées au fur à mesure est capital pour la crédibilité des conclusions finales (51).

- **Champs d'application**

Les essais adaptatifs pourraient être intéressants dans le cadre de techniques opérateur-dépendantes, en chirurgie, dans les études de techniques interventionnelles ou chirurgicales nécessitant une phase d'apprentissage ou de mise au point, et dans le cadre de l'évaluation des dispositifs médicaux.

7.2.2.3. Méthodes bayésiennes

Les méthodes bayésiennes peuvent être utilisées comme alternatives aux méthodes séquentielles classiques (77).

Les méthodes bayésiennes sont également utilisées pour la réalisation d'analyses intermédiaires. Ainsi, dans un essai contrôlé randomisé comparant stents à élution de paclitaxel et groupe témoin chez des patients emphysémateux sévères, l'analyse intermédiaire planifiée repose sur des méthodes bayésiennes (38).

Enfin, ces approches permettent de mieux gérer des critères de jugement multiples (78).

▪ **Contexte**

Dans les essais randomisés conventionnels, les données apportées par des études antérieures sont utilisées au moment de la conception de l'étude. Par la suite, seule l'information collectée au cours de l'essai est utilisée. Au contraire, l'approche bayésienne combine des informations existantes et les informations apportées par l'essai en cours (78).

Si on veut rester dans une approche bayésienne acceptable, les informations a priori doivent :

- rester non informatives et ne pas influencer le résultat de l'étude,
- ou utiliser des résultats d'autres études valides.

L'utilisation de méthodes bayésiennes dans le domaine des dispositifs médicaux est préconisée dans les récentes recommandations de la FDA (79).

▪ **Principe**

L'approche bayésienne permet d'utiliser dans l'analyse de l'essai des informations cliniques disponibles a priori. Elle consiste à utiliser ces informations déjà existantes, qui peuvent être fournies par la littérature. Certaines publications estiment possible d'utiliser également les avis d'experts en les modélisant. Ce point est discutable car il introduit une part importante d'arbitraire dans le résultat de l'étude.

L'information a priori et les résultats de l'essai sont considérés comme faisant partie d'un même ensemble de données.

Dans un essai clinique bayésien, l'incertitude sur un paramètre d'intérêt est décrite selon des probabilités, actualisées au cours de la collecte d'information dans l'essai. Les probabilités constituant la distribution a priori sont fondées sur les données des essais antérieurs (78) (80). A partir du théorème de Bayes, les probabilités a posteriori sont estimées à partir de l'essai en cours, conditionnellement aux probabilités a priori. Cette approche combine l'information antérieure avec les données fournies par l'essai (78). Ces modèles sont extrêmement dépendants des informations a priori.

L'estimation bayésienne ne fournit pas d'intervalles de confiance, mais des intervalles de crédibilités basés sur la distribution a posteriori (78). A la différence de l'approche dite « fréquentiste », il n'y a pas de tests statistiques mais des résultats avec l'intervalle de crédibilité à 95%. Ces méthodes offrent une certaine souplesse.

- **Avantages**

Les méthodes bayésiennes permettent de fournir la probabilité que le traitement soit efficace. De plus il permet un plus grand apport d'information pour la prise de décision. De manière similaire à une méta-analyse, l'utilisation de données antérieures valides permet d'augmenter l'information et la précision fournies par l'essai. Dans certains cas, ce type de méthode permet de réduire le nombre de sujets nécessaires, soit grâce à l'apport des informations antérieures, soit en utilisant des essais adaptatifs. Un autre avantage repose sur la plus grande flexibilité de ces techniques à des changements au cours d'un essai.

Ainsi, la méthode bayésienne permet d'utiliser toute l'information disponible (à la fois passée et actuelle) pour une technologie donnée. De plus, l'approche bayésienne est particulièrement intéressante dans les situations où le nombre de sujets est faible. Enfin, en utilisant des informations antérieures informatives, elle peut réduire les effectifs, et ainsi, réduire la durée et les ressources nécessaires d'un essai clinique (78,80).

- **Inconvénients**

L'inconvénient **majeur** est le risque de prise en compte d'une information arbitraire utilisée comme a priori, influençant totalement le résultat final.

La situation reviendrait à inclure dans une méta-analyse une importante étude biaisée en faveur de l'effet du traitement, compensant totalement les résultats d'une petite étude sans biais.

- **Contraintes**

Les analyses bayésiennes reposent sur la génération d'hypothèses et de simulations, notamment en ce qui concerne l'information a priori (fournie à partir de données antérieures ou de réunions d'experts), l'information a posteriori (qui sera obtenue à partir des données de l'essai), et le modèle mathématique utilisé pour combiner les deux. L'influence des covariables sur les événements ou les données manquantes doit également être envisagée, et des analyses de sensibilité planifiées (78).

Ce travail en amont est particulièrement important.

- **Champs d'application**

Alors que l'approche dite « fréquentiste » des essais classiques demeure encore privilégiée dans la littérature biomédicale, la publication d'essais faisant appel aux méthodes bayésiennes tend à se développer depuis quelques années (79).

7.3. Etudes observationnelles comparatives non randomisées

7.3.1. Type d'étude

Ce type d'étude ne peut être utilisé que dans les circonstances où il est impossible de réaliser un essai contrôlé randomisé. Le choix des études observationnelles doit demeurer l'exception. Les études observationnelles offrent un moins bon niveau de preuve que les essais randomisés.

7.3.1.1. Etude de cohorte complète⁸ ou essai basé sur la préférence des patients

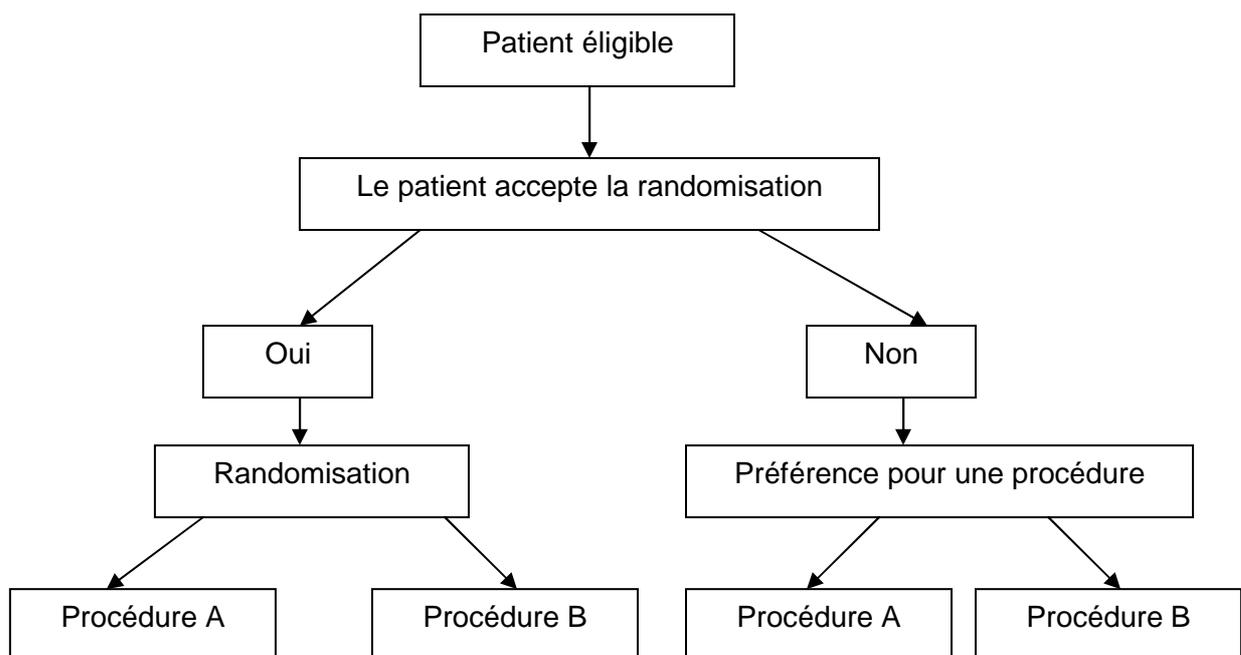
- Contexte

Dans les essais comparant un traitement chirurgical et un traitement non chirurgical, les patients peuvent percevoir le groupe avec traitement non chirurgical comme une intervention moins efficace, et refuser de participer à l'étude (81). L'exclusion des patients avec une forte préférence pour l'un ou l'autre des traitements peut altérer la validité externe des résultats. Les études de cohorte complète (ou « *comprehensive cohort design* », ou essais sur les préférences des patients ont été développés afin de surmonter ces obstacles.

- Principe

Tous les patients remplissant les critères d'éligibilité pour un essai sont recrutés, indépendamment de leur consentement à la randomisation (81). Lorsque le patient refuse la randomisation, il reçoit le traitement qu'il préfère (figure 11).

Figure 11 : Principe des études basées sur les préférences ou des « *comprehensive cohort study* » d'après (82)



⁸ « *comprehensive cohort study* »

- Avantages

Ce type d'étude permet d'améliorer le recrutement de patients participant à un essai clinique. De plus, il permet d'obtenir des informations sur l'acceptabilité des traitements, et de réduire le risque de sélectionner des patients peu représentatifs du fait des difficultés du recrutement. Enfin, il permet de concilier les avantages des essais randomisés en apportant l'opportunité de tester l'interaction entre la préférence et le critère de jugement (83).

- Inconvénients

L'affectation des patients à l'une ou l'autre des procédures selon leurs préférences implique les mêmes limites que celles constatées pour les études observationnelles : rien ne garantit que ces patients soient comparables sur les variables non mesurées.

Ce type d'étude est également sujet à des biais de sélection. La différence observée n'est pas due au traitement mais à d'autres facteurs non contrôlés (biais de confusion).

- Contraintes

La variable sur la randomisation est introduite dans le modèle comme covariable d'ajustement.

- Champs d'application

Ce type d'étude permet d'améliorer le recrutement de patients lorsqu'il est difficile d'obtenir leur consentement pour un essai randomisé classique. Il pourrait être particulièrement intéressant en chirurgie, dans les procédures interventionnelles et pour les dispositifs médicaux.

7.3.1.2. Etude observationnelle comparative prospective

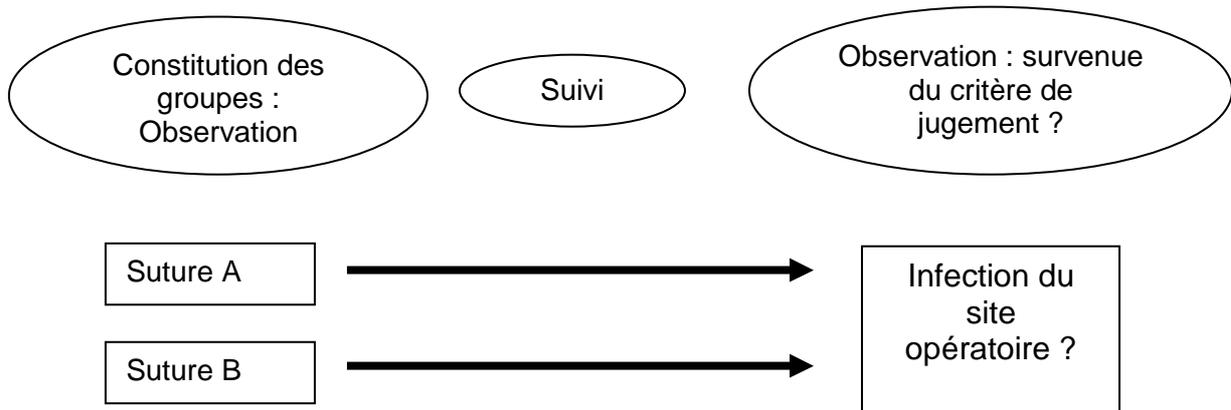
- Contexte

Les études observationnelles peuvent être utilisées lorsque les essais contrôlés randomisés ne peuvent pas être envisagés pour des raisons éthiques par exemple (62) Leur principale limite est la non-comparabilité des groupes, pouvant expliquer un éventuel effet traitement. La prise en compte des facteurs de confusion dans les études observationnelles est capitale (84).

- Principe

Les études comparatives non randomisées prospectives consistent à comparer l'incidence du critère de jugement (par exemple infection survenant après la chirurgie) entre un groupe recevant une procédure A (intervention ou dispositif médical, par exemple application d'un antiseptique X) et un groupe ne recevant pas la procédure A (soit il bénéficie d'une autre procédure, soit il ne reçoit rien). La répartition des patients dans chacun de ces groupes n'est pas établie par l'investigateur, c'est une donnée observée (figure 12). Le principe est le même que pour les études exposés-non exposés.

Figure 12 : Exemple d'étude observationnelle prospective comparant la survenue d'infections post-opératoires selon le type de suture utilisé



- Avantages

Une réalisation prospective permet de planifier le recueil de toutes les informations nécessaires. Elle peut permettre également de recueillir des informations précises sur la procédure (doses administrées dans un appareil d'imagerie par exemple, ou durée de la procédure). De plus, la mesure de l'incidence d'un événement, d'un risque relatif, et la prise en compte du délai de survenue de l'événement après la procédure sont possibles.

- Inconvénients

Comme toutes études observationnelles, les études comparatives prospectives sont sujettes aux biais de confusion. Rien ne garantit que les deux groupes soient comparables sur les autres variables, en particulier pour les variables non mesurées. De plus, le suivi peut être long ce qui majore le risque de perdus de vue.

- Contraintes

Le groupe de référence doit être comparable au groupe recevant la procédure d'intérêt sur les autres variables, et doit pouvoir être suivi avec les mêmes modalités que le groupe étudié. Lorsque le suivi est particulièrement long, des procédures doivent être mises en place pour limiter le risque de perdus de vue. La période d'observation doit être définie.

- Champs d'application

Actuellement il s'agit de manière large des études de cohorte ou exposés - non exposés. Lorsqu'un essai randomisé est inenvisageable, elles peuvent être utilisées pour la comparaison d'interventions très différentes comme des techniques chirurgicales, des dispositifs médicaux ou autres.

7.3.2. Stratégies d'analyse

7.3.2.1. Score de propension

- **Contexte**

Dans les études observationnelles, l'attribution d'un traitement ne se fait pas par randomisation et l'investigateur n'intervient pas dans l'attribution d'un traitement. L'investigateur se contente d'observer ce qui se passe entre deux groupes, l'un recevant la traitement A, et l'autre recevant le traitement B. La comparabilité des groupes n'est pas garantie. Le score de propension est une méthode statistique ayant pour but de tenter de réduire les biais de confusion liés aux études observationnelles (85).

- **Principe**

Le score de propension correspond à la probabilité conditionnelle pour un patient de recevoir le traitement A plutôt que le traitement B, en fonction de ses caractéristiques initiales (propension à recevoir un traitement en fonction de ses caractéristiques). L'objectif est d'équilibrer la distribution des covariables (par exemple âge, sexe, comorbidités, gravité de la maladie) entre les groupes, afin de neutraliser les facteurs de confusion (86,87). La comparaison entre groupe traité et non traité se fait au sein de la même classe de score de propension (85). Ainsi, les sujets traités et les témoins avec des scores de propension égaux peuvent être considérés comme similaires (84).

Une fois établi, le score de propension peut être utilisé comme variable d'appariement, comme variable de stratification, ou comme variable d'ajustement. Les méthodes de choix sont la stratification et l'appariement, par exemple en stratifiant l'analyse sur les quintiles du score de propension. Ainsi, au sein de chacune de ces classes, la plupart des biais liés aux facteurs de confusion mesurés disparaissent. L'appariement peut être plus complexe en raison de l'échelle continue du score de propension (84).

Exemple

Dans une étude visant à comparer la fréquence des thromboses veineuses profondes selon l'administration ou non d'héparine de bas poids moléculaire (HBPM), le score de propension correspondait à la probabilité, pour un patient, de recevoir une HBPM selon ses caractéristiques initiales. Une fois le score de propension déterminé, celui-ci était classé en quintiles de valeurs croissantes, les patients du 1^{er} quintile ayant la plus faible probabilité d'être sous HBPM et les patients du 5^{ème} quintile la plus forte (88). Le score de propension était utilisé comme variable de stratification. Les auteurs ont également utilisé le score de propension comme variable d'appariement : chaque patient traité était apparié au patient non traité avec le score d'appariement le plus proche et différent de moins de 0,1.

La méthode par score de propension fonctionne mieux sous 3 conditions (87) :

- lorsque l'événement étudié est rare
- lorsqu'il y a beaucoup de patients dans chaque groupe
- lorsque de nombreuses covariables ont été mesurées

- **Avantages**

Le score de propension permet de neutraliser les biais de confusion connus sur les variables mesurées en rendant les groupes comparables pour un score donné.

De plus, il permet d'ajuster simultanément sur des nombreuses variables déséquilibrées et ainsi de réduire les biais pour la comparaison entre 2 traitements (87). Ainsi, la stratification sur le score de propension réduit de plus de 90% les biais dus aux différences entre les caractéristiques observées (89).

- **Inconvénients**

L'ajustement ne peut être réalisé que sur les variables recueillies (87). La comparabilité des groupes n'est donc pas garantie pour les variables non mesurées (84). Cependant, la réalisation d'analyses de sensibilité peut permettre d'évaluer la robustesse des résultats (85,90).

L'analyse par score de propension n'est pas adaptée lorsque l'échantillon est petit (< 50). En effet, dans ce cas, certaines variables peuvent être déséquilibrées (87). Il en est de même en cas de déséquilibre sévère entre les groupes pour les variables à l'inclusion (87).

Enfin, le score de propension ne remplace pas un essai randomisé, et ne permet pas d'éliminer tous les biais de sélection (87).

- **Contraintes**

L'analyse par score de propension doit être planifiée à l'avance. Toutes les variables pertinentes à collecter doivent être spécifiées dans le protocole. De plus, les populations de patients entre les deux traitements doivent être comparables. Le nombre de sujets nécessaires doit être calculé avant la mise en route de l'étude en prenant en compte l'analyse par score de propension. Lors de l'analyse des résultats, la comparabilité de la distribution des groupes de score de propension doit être vérifiée. Enfin, les données manquantes doivent être estimées par des techniques spécifiques (imputation multiple par exemple) afin de ne pas exclure de patients, ce qui remettrait en cause la validité des résultats. Classiquement, une analyse de sensibilité est réalisée pour les variables non observées pour identifier les biais cachés (87).

- **Champ d'application**

L'analyse par score de propension présente un grand intérêt pour les études observationnelles lorsqu'un essai randomisé n'est pas réalisable compte-tenu des pratiques médicales, des préférences des patients, de l'organisation des soins ou des contraintes économiques (88). L'échantillon doit de plus être de taille suffisante. Cette stratégie d'analyse pourrait être utilisée pour les dispositifs médicaux, en chirurgie et pour les techniques interventionnelles.

7.3.2.2. Variables instrumentales.

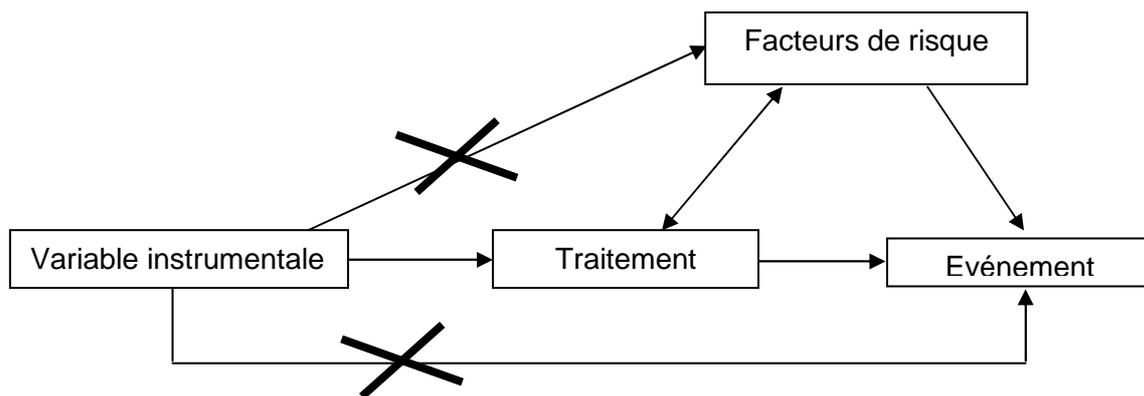
- **Contexte**

Les variables instrumentales sont largement utilisées en économétrie depuis des années, mais leur utilisation dans le domaine de la santé est plus récente (91). L'analyse par variable instrumentale permet de tirer avantage des données observationnelles pour estimer l'efficacité d'un traitement, même en présence de facteurs de risque non mesurés (92).

- **Principe**

Une variable instrumentale (comme la préférence du prescripteur) est une variable fortement associée à l'indication du traitement, mais non liée aux facteurs de risque de la survenue de l'événement et n'affectant pas directement le critère de jugement (92) (figure 13).

Figure 13 d'après (92).



Dans l'approche par variable instrumentale, on considère qu'un traitement est confondu par l'indication. Ainsi, les patients pourraient être sélectionnés pour recevoir l'un des deux traitements en raison de facteurs pronostiques, connus ou inconnus. Contrairement au score de propension qui a pour but d'ajuster sur les facteurs de confusion connus, l'approche par variable instrumentale vise également à éliminer les biais résultant de facteurs de confusion inconnus (93).

L'approche par variable instrumentale commence par l'identification de la variable instrumentale (94).

- **Avantages**

L'approche par variable instrumentale permet de réduire les biais de confusion dans les études observationnelles (94) et de prendre en compte les variables non mesurées (92,93).

- **Inconvénients**

L'estimation de l'effet du traitement est fondée sur des hypothèses fortes qui d'une part limitent son utilisation en pratique et qui sont d'autre part difficiles à vérifier.

De plus, l'effet du traitement pourrait ne pas être généralisable à une population de patients dont le statut thérapeutique n'a pas été déterminé par variable instrumentale. L'effet

traitement est estimé pour les patients « marginaux », ceux qui sont « compliants » ou « observants » (95). L'effet sur la population générale peut être différent.

Au total, lorsque la variation de la probabilité de recevoir un traitement particulier est faible entre des groupes de patients établis selon une variable instrumentale, les différences dans l'événement dues à l'utilisation différentielle de ce traitement pourraient être très petites, et, de ce fait, difficiles à évaluer (84).

Trouver des variables instrumentales valides est extrêmement difficile. En effet, la plupart des variables ayant un effet sur le traitement peuvent aussi avoir un effet direct sur l'événement (96). D'autre part, un bon estimateur de variable instrumentale repose sur des effectifs larges (96). En cas de petits effectifs, les estimations sont incorrectes. Enfin, l'interprétation est difficile, en particulier quand l'effet traitement est hétérogène.

- **Contraintes**

Une bonne variable instrumentale doit répondre à 3 hypothèses majeures (92) :

- avoir une relation forte avec l'indication du traitement ; cette estimation devrait être faite sans biais ;
- ne pas être corrélée aux facteurs de confusion (mesurés ou non) ;
- ne pas être reliée directement à l'événement, mais uniquement à travers l'effet du traitement.

- **Champ d'application**

Les variables instrumentales sont beaucoup utilisées dans le domaine de l'économétrie. En revanche, les contraintes inhérentes à ces variables les rendent difficilement envisageables dans le cadre de l'évaluation de dispositifs médicaux.

8. En pratique

Dans certaines circonstances, le développement clinique d'un nouveau DM ne permet pas de réaliser un essai contrôlé randomisé conventionnel. D'autres plans expérimentaux pourront être proposés afin de démontrer l'intérêt clinique du DM.

Leur choix peut être guidé par :

Les caractéristiques du dispositif médical :

- Une évolution technologique importante et rapide :
 - L'essai avec randomisation adaptative
 - Les méthodes bayésiennes
 - Eventuellement essais suiveurs ou « *tracker trial design* »
- Une population cible faible
 - Les méthodes bayésiennes
 - Les essais en cross-over à la condition que la pathologie étudiée soit stable et que le critère de jugement puisse se répéter
 - Les essais séquentiels
- Des événements indésirables potentiellement graves
 - Les essais séquentiels

L'acceptabilité médicale :

- La maîtrise de la technique influence le résultat ou les techniques sont très différentes
 - L'essai basé sur « l'expertise » ou « expertise-based randomized controlled trials »
 - Dans certaines circonstances l'essai en « cluster »

L'acceptabilité du patient

- Le comparateur est une technique invasive ou la technique est largement diffusée
 - Le plan expérimental de type « Zelen » ou « randomized consent trial »
 - Essais basés sur les préférences des patients ou étude observationnelle comparative si la randomisation est impossible

La figure 14 propose un arbre décisionnel pour orienter le choix du type d'étude. Le niveau de preuve respectif des différents types d'étude est illustré par la figure 15.

L'essai prospectif non randomisé est justifié uniquement lorsque tous les autres types d'essais ne peuvent pas être appliqués. La décision devra être justifiée à partir d'une argumentation scientifique explicite.

Figure 14 : Arbre décisionnel du type d'étude

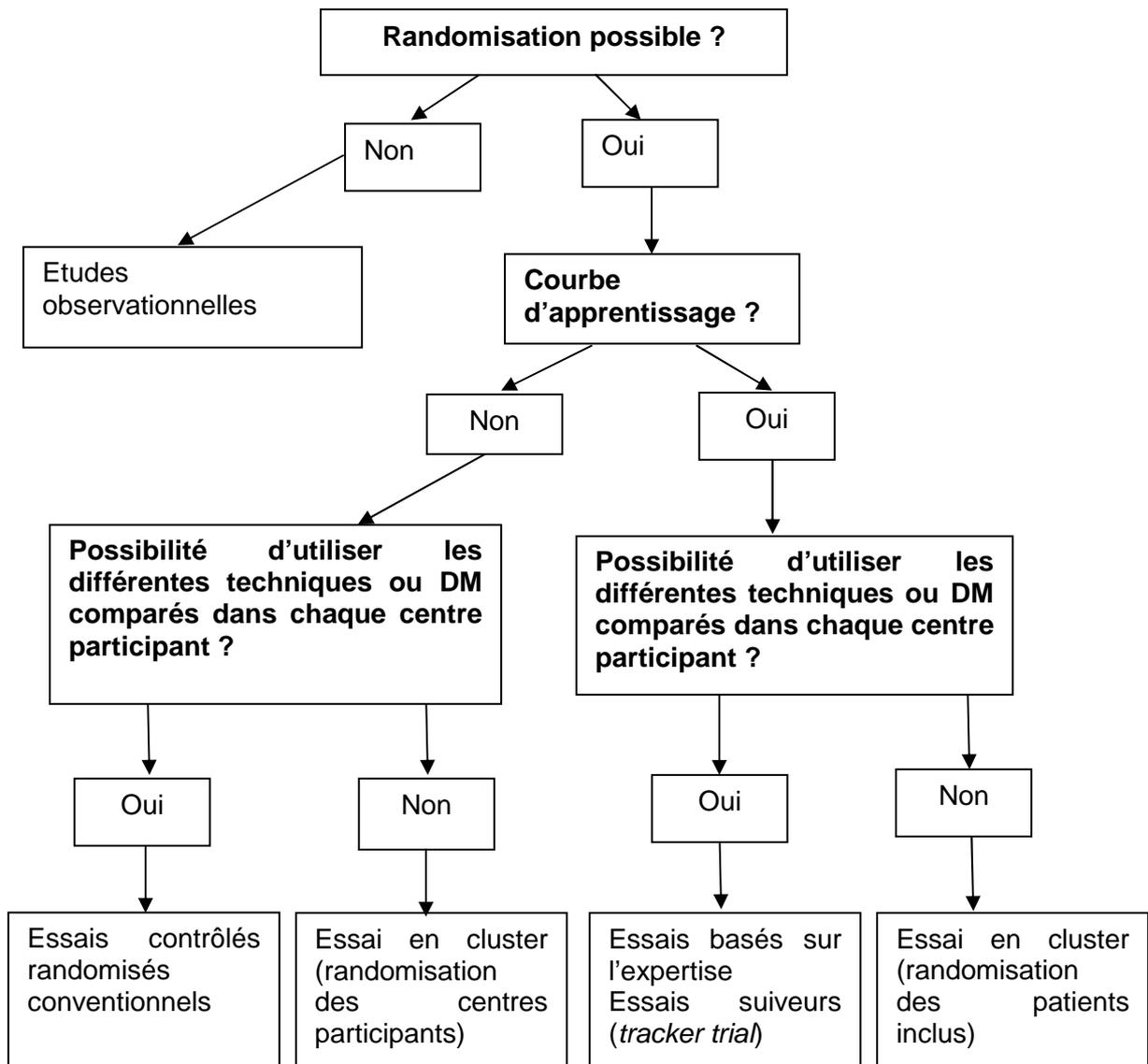
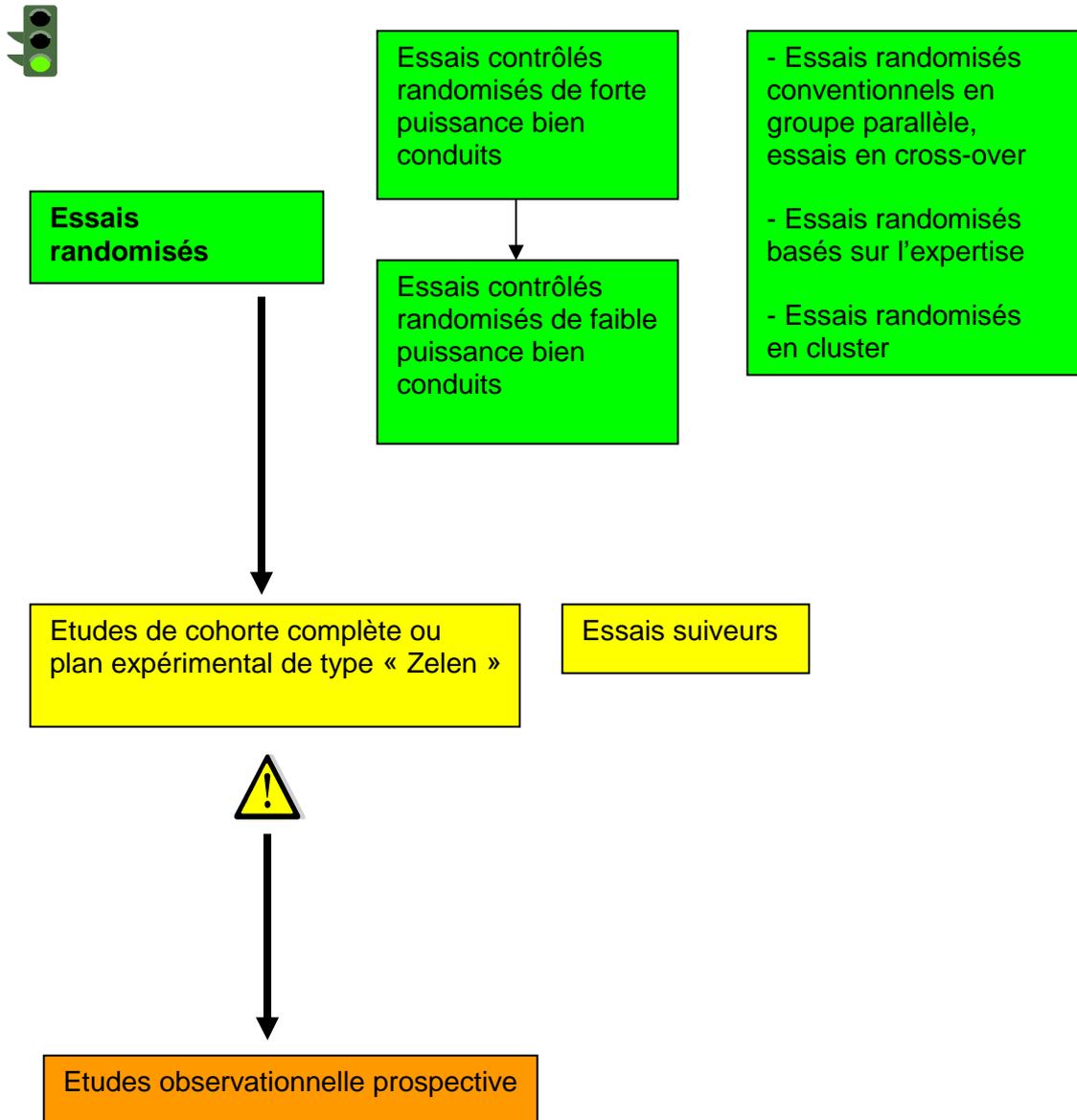


Figure 15 : Proposition de classement par niveau de preuve des types d'étude



Bibliographie

1. Agence nationale d'accréditation et d'évaluation en santé. Guide d'analyse de la littérature et gradation des recommandations. Saint-Denis la Plaine: ANAES; 2000. http://www.has-sante.fr/portail/upload/docs/application/pdf/analite_rat.pdf
2. Boutron I, Moher D, Tugwell P, Giraudeau B, Poiraudou S, Nizard R, *et al.* A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol* 2005;58(12):1233-40.
3. Centre fédérale d'expertise des soins en santé, Vinck,I, Neyt,M, Thiry,N, Louagie,M, Ghinet,D, *et al.* Procédure d'évaluation des dispositifs médicaux émergents. Bruxelles: Centre fédérale d'expertise des soins en santé (KCE); 2006.
4. EuroScan, Simpson,S, Hiller,J, Gutierrez-Ibarluzea,I, Kearney,B, Norderhaug,I, *et al.* A toolkit for the identification and assessment of new and emerging health technologies. Birmingham: EuroScan; 2009.
5. Bernard A. Le développement clinique des nouveaux dispositifs médicaux. e-mémoires de l'Académie Nationale de Chirurgie 2012;11(4):027-31.
6. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, *et al.* Challenges in evaluating surgical innovation. *Lancet* 2009;374(9695):1097-104.
7. Bernard A, Vicaut E. Quelles études pour évaluer des nouveaux dispositifs médicaux ? *J Chir* 2009;146(2):129-35.
8. Bernard A, Vicaut E. Dispositifs médicaux. Première partie : de quels types d'études avons-nous besoin ? *Médecine* 2008;4(10):464-7.
9. Bednarska E, Bryant D, Devereaux PJ. Orthopaedic surgeons prefer to participate in expertise-based randomized trials. *Clin Orthop Relat Res* 2008;466(7):1734-44.
10. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, *et al.* CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008;5(1).
11. Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA* 2003;290(8):1062-70.
12. Weil RJ. The future of surgical research. *PLoS Med* 2004;1(1):e13.
13. Freed CR, Greene PE, Breeze RE, Tsai WY, DuMouchel W, Kao R, *et al.* Transplantation of embryonic dopamine neurons for severe Parkinson's disease. *N Engl J Med* 2001;344(10):710-9.
14. Tamis-Holland JE, Palazzo A, Stebbins AL, Slater JN, Boland J, Ellis SG, *et al.* Benefits of direct angioplasty for women and men with acute myocardial infarction: results of the Global Use of Strategies to Open Occluded Arteries in Acute Coronary Syndromes Angioplasty (GUSTO II-B) Angioplasty Substudy. *Am Heart J* 2004;147(1):133-9.
15. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. Systematic reviews in health care. *BMJ* 2001;323(7303):42-6.
16. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol* 2004;57(6):543-50.
17. Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 2007;4(2).
18. Kinnaert P. La chirurgie factuelle : aspects éthiques et méthodologiques. *Rev Med Brux* 2006;27:451-8.
19. Kinnaert P. Placebo et effet placebo (deuxième partie) : aspects éthiques. *Rev Med Brux* 2007;28:39-44.
20. Ravaud,P. Evaluation des traitements non médicamenteux. 4^{ème} conférence d'épidémiologie clinique. 26-28 mai 2010. Paris: 2011.
21. Li H, Yue LQ. Statistical and regulatory issues in nonrandomized medical device clinical studies. *J Biopharm Stat* 2008;18(1):20-30.
22. Freeman TB, Vawter DE, Leaverton PE, Godbold JH, Hauser RA, Goetz CG, *et al.* Use of placebo surgery in controlled trials of a cellular-based therapy for Parkinson's disease. *N Engl J Med* 1999;341(13):988-92.
23. Macklin R. The ethical problems with sham surgery in clinical research. *N Engl J Med*

- 1999;341(13):992-6.
24. Macklin R. Placebo surgery in trials of therapy for Parkinson's disease. *N Engl J Med* 2000;342(5):355.
25. Master Z, McLeod M, Mendez I. Benefits, risks and ethical considerations in translation of stem cell research to clinical applications in Parkinson's disease. *J Med Ethics* 2007;33(3):169-73.
26. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, *et al.* Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med* 2006;3(10):e425.
27. Rudicel S, Esdaile J. The randomized clinical trial in orthopaedics: obligation or option? *J Bone Joint Surg Am* 1985;67(8):1284-93.
28. Campbell G. Statistics in the world of medical devices: the contrast with pharmaceuticals. *J Biopharm Stat* 2008;18(1):4-19.
29. Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *Lancet* 2003;361(9351):13-20.
30. Stone GW, Grines CL, O'Neill WW. Primary coronary angioplasty versus thrombolysis. *N Engl J Med* 1997;337(16):1168-9.
31. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med* 2002;137(6):511-20.
32. Zelen M. A new design for randomized clinical trials. *N Engl J Med* 1979;300(22):1242-5.
33. Blondel B, Goffinet F, Bréart. *Evaluation des soins en obstétrique*. Paris: Masson; 2001.
34. Zelen M. Innovations in the design of clinical trials in breast cancer. *Breast Cancer Res Treat* 1983;3(2):137-42.
35. Zelen M. Strategy and alternate randomized designs in cancer clinical trials. *Cancer Treat Rep* 1982;66(5):1095-100.
36. Zelen M. Randomized consent designs for clinical trials: an update. *Stat Med* 1990;9(6):645-56.
37. Horwitz RI, Feinstein AR. Advantages and drawbacks of the Zelen design for randomized clinical trials. *J Clin Pharmacol* 1980;20(7):425-7.
38. Shah PL, Slebos DJ, Cardoso PFG, Cetti EJ, Sybrecht GW, Cooper JD. Design of the exhale airway stents for emphysema (EASE) trial: an endoscopic procedure for reducing hyperinflation. *BMC Pulm Med* 2011;11:1.
39. Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, *et al.* Need for expertise based randomised controlled trials. *BMJ* 2005;330(7482):88-91.
40. Mastracci TM, Clase CM, Devereaux PJ, Cinà CS. Open versus endovascular repair of abdominal aortic aneurysm: a survey of Canadian vascular surgeons. *J Can Surg* 2008;51(2):142-8.
41. DeTurris SV, Cacchione RN, Mungara A, Pecoraro A, Ferzli GS. Laparoscopic herniorrhaphy: beyond the learning curve. *J Am Coll Surg* 2002;194(1):65-73.
42. Menon VS, Manson JMCK, Baxter JN. Laparoscopic fundoplication: learning curve and patient satisfaction. *Ann R Coll Surg Engl* 2003;85(1):10-3.
43. Lobato AC, Rodriguez-Lopez J, Diethrich EB. Learning curve for endovascular abdominal aortic aneurysm repair: evaluation of a 277-patient single-center experience. *J Endovasc Ther* 2002;9(3):262-8.
44. Lilford RJ, Braunholtz DA, Greenhalgh R, Edwards SJL. Trials and fast changing technologies: the case for tracker studies. *BMJ* 2000;320(7226):43-6.
45. Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomised trials. *BMC Med Res Methodol* 2005;5:10.
46. Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. *PLoS Med* 2009;6(5).
47. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000;9(2):81-94.
48. Puffer S, Torgerson DJ, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ* 2003;327(7418):785-9.
49. Elbourne DR, Altman DG, Higgins JPT, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol* 2002;31(1):140-9.
50. Hills M, Armitage P. The two-period cross-over clinical trial. *Br J Clin Pharmacol*

1979;8(1):7-20.

51. van der Lee JH, Wesseling J, Tanck MWT, Offringa M. Efficient ways exist to obtain the optimal sample size in clinical trials in rare diseases. *J Clin Epidemiol* 2008;61(4):324-30.

52. Wassmer G. Basic concepts of group sequential and adaptative group sequential test procedures. *Stat Pap* 2000;41:253-79.

53. Armitage P. *Sequential medical trials*. 2nd ed. New York: Wiley; 1975.

54. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977;64(2):191-9.

55. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979;35(3):549-56.

56. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, *et al*. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976;34(6):585-612.

57. Falissard B, Lellouch J. Some extensions to a new approach for interim analysis in clinical trials. *Stat Med* 1991;10(6):949-55.

58. Falissard B, Lellouch J. A new procedure for group sequential analysis in clinical trials. *Biometrics* 1992;48(2):373-88.

59. DeMets DL, Ware JH. Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika* 1980;67(3):651-60.

60. DeMets DL, Ware JH. Asymmetric group sequential boundaries for monitoring clinical trials. *Biometrika* 1982;69(3):661-3.

61. Whitehead J, Stratton I. Group sequential clinical trials with triangular continuation regions. *Biometrics* 1983;39(1):227-36.

62. Clayton DG. Ethically optimised designs. *Br J Clin Pharmacol* 1982;13(4):469-80.

63. Benichou J, Chastang C. A simulation study of three sequential methods for the comparison of two treatment groups when the response criterion is censored. *Stat Med* 1986;5(4):375-85.

64. Benichou J, Chastang CK. Mise en œuvre de méthodes séquentielles dans les essais thérapeutiques dont le critère de jugement est censuré. *Rev Epidemiol Santé Publ* 1986;34:196-208.

65. Pocock SJ. Interim analyses for randomized clinical trials: the group sequential approach. *Biometrics* 1982;38(1):153-62.

66. Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials* 1989;10(4 Suppl):209S-21S.

67. Lan GKK, Simon R, Halperin M. Stochastically curtailed tests in long-term clinical trials. *Sequential Analysis* 1982;1(3):207-19.

68. Sebillé V, Bellissant E. Comparison of four sequential methods allowing for early stopping of comparative clinical trials. *Clin Sci* 2000;98(5):569-78.

69. Benichou J, Chastang CI. Analyse séquentielle des essais thérapeutiques randomisés dont le critère de jugement est la survie. Utilisation du test triangulaire. *Thérapie* 1987;42(3):295-9.

70. Coad DS, Ivanova A. The use of the triangular test with response-adaptive treatment allocation. *Stat Med* 2005;24(10):1483-93.

71. Bauer P, Brannath W. The advantages and disadvantages of adaptive designs for clinical trials. *Drug Discov Today* 2004;9(8):351-7.

72. Weinstein MC. Allocation of subjects in medical experiments. *N Engl J Med* 1974;291(24):1278-85.

73. Day NE. A comparison of some sequential designs. *Biometrika* 1969;56:301-11.

74. Lellouch J, Schwartz D. L'essai thérapeutique: Ethique individuelle ou éthique collective. *Rev Inst Int Stat* 1971;39:127-36.

75. Simon R. Adaptive treatment assignment methods and clinical trials. *Biometrics* 1977;33(4):743-9.

76. Hung HMJ. Considerations in adapting clinical trial design. *J Formos Med Assoc* 2008;107(12 Suppl):S14-S18.

77. Freedman LS, Spiegelhalter DJ. Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Control Clin Trials* 1989;10(4):357-67.

78. Center for Devices and Radiological Health, Center for Biologics Evaluation and Research. Guidance for industry and FDA staff. Guidance for the use of bayesian statistics in medical device clinical trial. Silver Spring, Rockville: CDRH, CBER; 2006.

<http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf>

79. Campbell G. Bayesian statistics in medical devices: innovation sparked by the FDA. *J Biopharm Stat* 2011;21(5):871-87.
80. Pibouleau L, Chevret S. Bayesian statistical method was underused despite its advantages in the assessment of implantable medical devices. *J Clin Epidemiol* 2011;64(3):270-9.
81. Kearney RS, Achten J, Parsons NR, Costa ML. The comprehensive cohort model in a pilot trial in orthopaedic trauma. *BMC Med Res Methodol* 2011;11:39.
82. Olschewski M, Scheurlen H. Comprehensive Cohort Study: an alternative to randomized consent design in a breast preservation trial. *Methods Inform Med* 1985;24(3):131-4.
83. Torgerson DJ, Sibbald B. Understanding controlled trials. What is a patient preference trial? *BMJ* 1998;316(7128):360.
84. Klungel OH, Martens EP, Psaty BM, Grobbee DE, Sullivan SD, Stricker BH, *et al.* Methods to assess intended effects of drug treatment in observational studies are reviewed. *J Clin Epidemiol* 2004;57(12):1223-31.
85. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984;79(387):516-24.
86. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41-55.
87. Yue LQ. Statistical and regulatory issues with the application of propensity score analysis to nonrandomized medical device clinical studies. *J Biopharm Stat* 2007;17(1):1-13.
88. Labarère J, Bosson JL, Francois P, Fine MJ. L'analyse par score de propension. Exemple

d'application à une étude observationnelle sur la prophylaxie de la maladie thromboembolique veineuse. *Rev Med Interne* 2008;29(3):255-8.

89. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA* 2007;297(3):278-85.
90. Rosenbaum PR. Discussing hidden bias in observational studies. *Ann Intern Med* 1991;115(11):901-5.
91. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;29(4):722-9.
92. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data : a systematic review. *J Clin Epidemiol* 2011;64(6):687-700.
93. Schmoor C, Gall C, Stampf S, Graf E. Correction of confounding bias in non-randomized studies by appropriate weighting. *Biom J* 2011;53(2):369-87.
94. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010;19(6):537-54.
95. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol* 2009;62(12):1226-32.
96. Baser O. Too much ado about instrumental variable approach: is the cure worse than the disease? *Value Health* 2009;12(8):1201-9.

Annexe 1 : Glossaire

Validité externe d'une étude :

Il s'agit de la possibilité de généraliser les résultats d'une étude afin de les appliquer à la population générale

Validité interne d'une étude :

Rend compte du fait que les résultats obtenus sont corrects pour la population étudiée ; la méthodologie et l'existence de biais déterminent la validité interne.

Les différents biais

- Le biais de sélection survient lorsque les deux groupes de l'essai ne sont pas comparables. Une différence entre les deux groupes peut alors apparaître en dehors de tout effet traitement. La randomisation a pour but d'éviter le biais de sélection en créant, en moyenne, deux groupes comparables.
- Le biais de confusion est le biais entraîné par l'absence de prise en considération des facteurs de confusion.
- Un biais de suivi (ou de performance ou d'exécution) survient lorsque les deux groupes ne sont pas suivis de la même manière au cours de l'essai et les soins appliqués sont différents. La comparabilité initiale est alors détruite et une différence peut apparaître en dehors de tout effet traitement.
- Le biais d'attrition (ou d'exclusion) survient lorsque les « sorties » d'étude (perdus de vue et données manquantes) sont différentes dans les deux groupes thérapeutiques.
- Le biais d'évaluation (ou de mesure ou de détection) survient quand la mesure du critère de jugement n'est pas réalisée de la même manière dans les deux groupes. Le double insu limite le risque de biais d'évaluation.
- Biais liés à l'analyse en intention de traiter : survient lorsque les patients ne sont pas analysés dans leur groupe de randomisation initial. Des exclusions secondaires sont susceptibles de biaiser le résultat, principalement en détruisant la comparabilité initiale des groupes et ce d'autant que les exclusions sont potentiellement liées à l'effet du traitement.

Risques alpha et beta : risques de conclure à tort en fonction de fluctuations aléatoires, susceptibles de conduire à des conclusions erronées à partir de l'observation ; le risque alpha est le risque de conclure à l'existence d'une différence qui n'existe pas ; le risque beta est le risque de conclure à tort à l'absence d'une différence, lorsqu'elle existe.

Intervalle de confiance (IC) : généralement fixé à 95%, il s'agit de l'intervalle de valeurs ayant 95% de chance de contenir la vraie valeur du paramètre estimé ; cet intervalle permet de visualiser l'incertitude de l'estimation.

Intervalle de crédibilité : dans l'approche bayésienne, cet intervalle, déduit de la distribution *a posteriori*, rend compte de la confiance que l'on peut avoir au sujet de la valeur du paramètre concerné, c'est à dire la probabilité à 95% que sa valeur réelle soit comprise entre les bornes de l'intervalle.

Analyse en intention de traiter (ITT) : consiste à analyser les patients dans leur groupe de randomisation, quel que soit le traitement reçu.

Analyse en per-protocole (PP) : consiste à analyser les patients selon le traitement qu'ils ont reçu, et non pas selon leur groupe de randomisation

Facteur de confusion : facteur ayant une association avec le traitement et qui peut influencer également le résultat ; il peut affaiblir ou renforcer une association entre l'exposition et les résultats observés.

Censure non aléatoire : donnée manquante induite par un patient perdu de vue (écart au protocole non du au hasard mais en rapport avec le traitement administré).

Score de propension : propension à recevoir un traitement en fonction des caractéristiques du patient.

Annexe 2 : Tableaux récapitulatifs

Tableau 5 : Stratégies d'analyse dans les études observationnelles

Type d'étude	Principes	Avantages	Inconvénients
Score de propension	Utilisation d'un score résumant les caractéristiques des covariables observées.	Permet de réduire le biais de confusion.	Ne prend en compte que les variables observées
Variable instrumentale	Utilisation d'une variable fortement associée à l'indication du traitement, mais non liée aux facteurs de risque de survenue de l'événement et n'affectant pas le critère de jugement	<ul style="list-style-type: none"> - Prise en compte des facteurs de confusion mesurés et non mesurés - Permet d'estimer l'efficacité d'un traitement même en présence de facteurs de risque non mesurés 	<ul style="list-style-type: none"> - Utilisation limitée en pratique - Possibilités de généralisation limitées - Variable instrumentale difficile à trouver - Estimation incorrecte si effectifs faibles

Tableau 6 : Plans d'analyse particuliers pour les essais randomisés

Type d'analyse	Principes	Avantages	Inconvénients
Essai séquentiel	Analyse intermédiaire permettant l'arrêt précoce d'un essai. Les résultats de l'essai concernant les patients inclus sont analysés avant la randomisation de nouveaux patients	<ul style="list-style-type: none"> - pas d'inflation du risque alpha - conservation de la puissance - minimisation du nombre de sujets nécessaires - évite les problèmes éthiques sur la poursuite de l'étude si la différence entre traitement est plus importante que celle attendue. 	<ul style="list-style-type: none"> - ne permet de comparer que 2 traitements - problème de l'évaluation des critères de jugement secondaire ou des effets indésirables pour lesquels on risque de manquer de puissance - frontière fixée a priori et figée - surestimation de la taille de l'effet - non représentativité des essais de faible effectif
Essai adaptatif	Analyse intermédiaire permettant l'arrêt précoce d'un essai. Le schéma de l'essai peut être changé sur la base des résultats observés dans l'analyse intermédiaire.	<ul style="list-style-type: none"> - pas d'inflation du risque alpha - conservation de la puissance - réduction du nombre de sujets nécessaires - maximise le nombre total de patients traités correctement en changeant le plan d'essai en fonction des résultats des analyses intermédiaires : intérêt sur le plan éthique car plus de patients correctement traités - grande flexibilité 	<ul style="list-style-type: none"> - difficultés pratiques - validité interne remise en cause
Méthodes bayésiennes	Utilisation de données a priori et intégration de données collectées au cours de l'essai	<ul style="list-style-type: none"> - Plus flexible - Diminution du nombre de sujets nécessaires - Réduction de la durée - Utilisation de toute l'information à priori disponible 	<ul style="list-style-type: none"> - Le résultat dépend parfois beaucoup plus de l'a priori considéré plutôt que des données obtenues - Pas de test statistique - Pas d'intervalle de confiance (mais équivalent : intervalle de crédibilité)
Analyse en biais maximal	Lorsque l'échec avec une procédure conduit à réaliser l'autre procédure comparée, avec une issue favorable, le patient sera considéré en échec dans son groupe de randomisation	<ul style="list-style-type: none"> - Permet de ne pas considérer à tort un succès qui serait en fait dû à la technique de sauvetage 	Perte de contraste : réduction des différences observées entre les groupes

Tableau 7 : Types d'essais randomisés non conventionnels

Type d'essai	Principes	Avantages	Inconvénients
"Expertise-based randomized controlled trials"	Randomisation des patients auprès d'un praticien qui réalise la procédure pour laquelle il est « expert »	<ul style="list-style-type: none"> - diminution du risque d'écart au protocole différentiel - limitation des biais de suivi différentiel - amélioration de la validité interne - Diminution du risque de biais différentiel lié à l'expertise - Meilleure acceptabilité - Meilleure faisabilité (réduction des délais pour débiter l'essai) - Moins de problème éthique 	<ul style="list-style-type: none"> - Si différence : liée à la technique ou au chirurgien maîtrisant la technique ? - Délai d'inclusion plus long - Liste d'attente
Zelen	Randomisation avant le consentement	<ul style="list-style-type: none"> - Moins de retrait de consentement dans le groupe contrôle - Amélioration de la faisabilité du recrutement <ul style="list-style-type: none"> - Pas d'altération de la relation médecin malade 	<ul style="list-style-type: none"> - Biais de sélection - Problèmes éthiques - Aveugle impossible - Dilution de l'effet du traitement - Non utilisable dans les essais contre placebo
« Tracker trial design »	Les modifications au cours de l'essai sont autorisées, enregistrées et prises en compte dans l'analyse statistique. Les variations dans le schéma d'étude sont aussi autorisées.	<ul style="list-style-type: none"> - évaluation précoce - respect du principe d'équilibre - maximisation des données recueillies 	<ul style="list-style-type: none"> - Très difficile en pratique en termes d'analyse et de logistique - Problème du degré d'expérience différent des opérateurs
Essai en « cluster »	Randomisation de groupes d'individus plutôt que des individus directement	<ul style="list-style-type: none"> - Prévient le biais de contamination - Plus simple sur le plan logistique 	<ul style="list-style-type: none"> - Comparabilité des groupes non systématiquement garantie - Biais de recrutement différentiel - Augmentation du nombre de sujets nécessaires - Problèmes éthiques pour le consentement
Cross-over	Chaque sujet reçoit alternativement les différents traitements après une période de sevrage. L'ordre d'administration est tiré au sort.	<ul style="list-style-type: none"> - Permet d'augmenter la puissance à nombre de sujet égale - Meilleure précision 	<ul style="list-style-type: none"> - Nécessite une pathologie stable - Les traitements ne doivent pas avoir d'effet définitif

Tableau 8 : Procédures permettant d'accroître le recrutement de patients

Type d'essai	Principes	Avantages	Inconvénients
Zelen	Randomisation avant le consentement	<ul style="list-style-type: none"> - Moins de retrait de consentement dans le groupe contrôle - Amélioration de la faisabilité du recrutement - Pas d'altération de la relation médecin malade 	<ul style="list-style-type: none"> - Biais de sélection - Problèmes éthiques - Aveugle impossible - Dilution de l'effet du traitement - Pas utilisable dans les essais contre placebo
Etude de cohorte complète	<p>Tous les patients éligibles se voient proposer de participer à l'essai.</p> <p>S'ils acceptent la randomisation, ils sont randomisés pour l'une ou l'autre des procédures.</p> <p>S'ils refusent la randomisation, ils se voient demander quelle procédure ils préfèrent recevoir</p>	<ul style="list-style-type: none"> - Permet d'augmenter le recrutement des sujets - Permet de recueillir des informations sur l'acceptabilité des traitements - Permet de limiter les défauts de représentativité observés dans les essais cliniques (population très sélectionnée, pouvant être éloignée de celle observée en routine) 	<p>Les bras correspondant aux préférences des patients correspondent à une étude observationnelle, et sont donc sujets aux biais de confusion, avec une comparabilité des groupes qui n'est plus garantie.</p>



Toutes les publications de l'HAS sont téléchargeables sur
www.has-sante.fr

n° ISBN : 978-2-11-138103-2