



HAUTE AUTORITÉ DE SANTÉ

ÉVALUER

LES TECHNOLOGIES DE SANTÉ

GUIDE
METHODOLOGIQUE

Méthodologie pour le développement clinique des dispositifs médicaux

Validé par la CNEDiMTS le 8 juin 2021

Mis à jour en juin 2021

Descriptif de la publication

Titre	Méthodologie pour le développement clinique des dispositifs médicaux
Méthode de travail	<p>Guide méthodologique élaboré à partir d'une revue de la littérature, d'un état des lieux des recommandations françaises et internationales et de l'avis d'experts réunis dans un groupe de pilotage et un groupe de lecture.</p> <p>La méthode de travail comporte les différentes étapes suivantes :</p> <ul style="list-style-type: none"> – définition du projet par le groupe de pilotage ; – recherche bibliographique : veille de la recherche effectuée en 2013 et nouvelle recherche sur les thématiques complémentaires (méthodologies spécifiquement développées pour les populations cibles de petit effectif, méthodes <i>in silico</i>, études construites sur des bases de données médico-administratives) ; – sélection et analyse bibliographiques ; – rédaction du guide : actualisation du guide de 2013 et rédaction des nouvelles parties ; – relecture du guide par le groupe de lecture ; – finalisation du guide par le groupe de pilotage en tenant compte des commentaires du groupe de lecture ; – examen en CNEDiMTS en vue de la validation de la version finale du guide méthodologique actualisé.
Objectif(s)	Ce guide méthodologique vise à soutenir et accompagner le développement des dispositifs médicaux. Il vise à donner des points de repère pratiques sur les aspects méthodologiques pour optimiser le niveau de preuve des différents types d'études et renforcer la confiance dans leurs résultats.
Cibles concernées	Le guide méthodologique actualisé est destiné aux industriels, aux structures de recherche, aux porteurs de projet et aux professionnels de santé impliqués dans le développement clinique d'un dispositif médical et qui envisagent de déposer une demande d'admission au remboursement du dispositif concerné auprès de la CNEDiMTS. Les patients sont également concernés, puisque ce guide est destiné à favoriser le recueil de la preuve, quel que soit le critère utilisé (critère clinique, critère de qualité de vie, critère organisationnel).
Demandeur	Autosaisine
Promoteur(s)	Haute Autorité de santé (HAS)
Pilotage du projet	Service évaluation des dispositifs (SED)
Recherche documentaire	Service documentaire
Auteurs	<p>Groupe de pilotage : équipe projet du SED pour la HAS : Aurélie GAREL PACULL, Emmanuelle FOUTEAU, Corinne COLLIGNON ; membres du service documentation : Sophie DESPEYROUX, Juliette CHAZARENG ; membres de la CNEDiMTS : Dominique COSTAGLIOLA, Pascal SELLIER, Sarah ZOHAR.</p> <p>Groupe de lecture : Alain BERNARD, Michel CUCHERAT, Pascal GIRAUX, SNITEM.</p> <p>Avec la participation d'Éric DARVOY (schémas et illustrations), Laura DRUART (charte graphique), Isabelle DUBRAIL et Amara HRUSTIC (secrétariat).</p> <p>Rappel des participants à la première version de 2013 : groupe projet : Isabelle FOURNEL (chargée de projet) ; SED : Michel VANEAU ; membres du service documentation : Aurélien DANCOISNE, Laurence FRIGERE ; secrétariat : Sandrine BOUVET, Fadela CHEBILI. Groupe de travail : Alain BERNARD, Michel CUCHERAT, Isabelle FOURNEL, Hubert GALMICHE, Pascal GIRAUX, Bernard GUILLOT, Jacques</p>

MACHECOURT, Patrick MAISON. Groupe de lecture : Michèle MORIN-SUROCCA, Patrice NONY, Françoise ROUDOT-THORAVALE, Sophie STAMENKOVIC.

La HAS tient à remercier l'ensemble des participants cités ci-dessus.

Conflits d'intérêts	Les membres du groupe de pilotage ont communiqué leurs déclarations publiques d'intérêts à la HAS. Elles sont consultables sur le site https://dpi.sante.gouv.fr . Elles ont été analysées selon la grille d'analyse du guide des déclarations d'intérêts et de gestion des conflits d'intérêts de la HAS. Les intérêts déclarés par les membres du groupe de pilotage ont été considérés comme étant compatibles avec leur participation à ce travail.
Validation	Version du 8 juin 2021
Actualisation	

Sommaire

1. Introduction	6
2. Objectif	8
3. Contexte	9
3.1. L'étude contrôlée randomisée en double aveugle	9
3.2. Phases du développement clinique	9
4. Particularités méthodologiques pour le développement clinique des dispositifs médicaux	11
4.1. Choix du moment de l'évaluation	11
4.2. Population éligible et recrutement	11
4.3. Acceptabilité	11
4.4. Randomisation	12
4.5. Aveugle	12
4.6. Choix du groupe de référence ou comparateur	13
4.7. Facteurs liés à l'expérience des opérateurs	15
4.7.1. Courbe d'apprentissage	15
4.7.2. Volume des actes	15
4.8. Facteurs d'ajustement	16
4.9. Types d'analyse	16
4.9.1. Analyse en intention de traiter	16
4.9.2. Analyse en per-protocole	16
5. Études expérimentales	17
5.1. Obtenir l'aveugle ou compenser son absence	17
5.2. Zelen ou <i>randomised consent design trial</i>	18
5.3. Essais fondés sur l'expertise (<i>expertise-based randomised controlled trials</i>)	22
5.4. <i>Tracker trial design</i> ou essais suiveurs	25
5.5. Essais en <i>cluster</i>	26
5.6. Essais en <i>cross-over</i>	29
5.7. Études expérimentales en cas unique (SCED) [Nouveauté]	32
5.8. Études <i>in silico</i> [Nouveauté]	35
5.9. Types d'analyse	36
5.9.1. Méthodes bayésiennes [Nouveauté]	36
5.9.2. Essais séquentiels	38
5.9.3. Essais adaptatifs	42
6. Études en vie réelle ou études observationnelles [Nouveauté]	44

6.1. Principes généraux [Nouveauté]	44
6.2. Études observationnelles comparatives spécifiques à une collecte prospective des données	47
6.3. Cohortes complètes selon la préférence patient	50
6.4. Études à partir de données existantes [Nouveauté]	52
6.5. Comparaisons indirectes à partir d'études mono-bras [Nouveauté]	54
6.6. Types d'analyse	55
6.6.1. Méthodes d'ajustement et score de propension [Nouveauté]	55
6.6.2. Variables instrumentales	58
7. En pratique	60
Table des annexes	62
Références bibliographiques	72
Abréviations et acronymes	82
Glossaire	83

Ce document ainsi que sa référence bibliographique sont téléchargeables sur www.has-sante.fr 

Haute Autorité de santé – Service communication et information

5, avenue du Stade de France – 93218 Saint-Denis La Plaine Cedex. Tél. : +33 (0)1 55 93 70 00

© Haute Autorité de santé – juin 2021 – ISBN : 978-2-11-162650-8

1. Introduction

La commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (CNE-DiMITS) est la commission de la Haute Autorité de santé (HAS) qui évalue notamment les dispositifs médicaux (DM) et autres produits de santé en vue de leur remboursement par l'Assurance maladie (article L. 165-1 du Code de la sécurité sociale). Son rôle est de donner au décideur un avis consultatif recommandant ou non la prise en charge des DM (inscription sur la liste des produits et prestations remboursables ou LPPR, modifications des conditions d'inscription, renouvellement), de contribuer à déterminer les conditions de leur bon usage ainsi que leur place dans la stratégie thérapeutique, diagnostique ou de compensation du handicap.

La mission d'évaluation scientifique de la CNEDiMITS n'intervient que lorsque le marquage CE a été obtenu (article R. 165-4 du Code de la sécurité sociale). Son évaluation est donc complémentaire de celle mise en œuvre pour le marquage CE : au-delà de la démonstration des performances et de la sécurité, elle s'attache à évaluer l'intérêt du DM pour le patient et pour la santé publique (service attendu), ainsi que sa place dans l'arsenal disponible en France (amélioration du service attendu).

L'évaluation de l'intérêt des DM doit reposer sur des études cliniques probantes. Les avis de la CNE-DiMITS prennent en compte le contexte scientifique et médical existant au moment de l'évaluation ainsi que les données cliniques disponibles, principalement issues des essais cliniques lors de la demande d'inscription au remboursement. Si les études cliniques contrôlées randomisées sont généralement considérées comme la méthodologie de référence pour démontrer l'efficacité d'un produit de santé, conformément aux principes de l'*Evidence Based Medicine*, elles sont parfois difficiles à mettre en œuvre pour les dispositifs médicaux. Les spécificités du secteur, désormais bien identifiées telles que le rythme rapide de développement, le caractère opérateur dépendant ou lié à l'organisation des soins, les populations cibles parfois très restreintes, impliquent le développement de modes de preuve adaptés. Enfin, l'arrivée de technologies embarquant de l'intelligence artificielle, le développement de l'accès aux données de santé en conditions réelles d'utilisation, l'évolution de plus en plus rapide des technologies et le dynamisme du secteur du dispositif médical imposent d'adapter les méthodes d'évaluation conjuguant une évaluation basée sur des méthodologies robustes adaptées au contexte de la technologie évaluée.

C'est pourquoi, dans un contexte en **évolution continue**, la commission a souhaité mettre à jour son guide publié en 2013 et initialement intitulé « Choix méthodologiques pour le développement clinique des dispositifs médicaux ». Pour rendre plus lisibles les objectifs de ce guide, qui est une boîte à outils pour permettre aux entreprises de construire leur plan de développement, la commission l'a renommé « Méthodologies pour le développement clinique des dispositifs médicaux ». Ce guide actualisé vient en complément de l'accompagnement mis en place depuis plusieurs années par la HAS pour les entreprises qui le souhaitent. En particulier pour les plans de développement complexes, elles peuvent solliciter des **rencontres précoces** pour échanger sur une étude clinique avant sa mise en œuvre.

Ce guide est plus généralement destiné aux industriels, aux structures de recherche, aux CRO, aux porteurs de projet et aux professionnels de santé impliqués dans le développement clinique d'un dispositif médical ou d'une technologie de santé¹ et qui envisagent de déposer une demande d'admission au remboursement du dispositif concerné auprès de la CNEDiMITS. Les patients sont également concernés, puisque ce guide est destiné à favoriser le recueil de la preuve, quel que soit le critère utilisé (critère clinique, critère de qualité de vie, critère organisationnel). Il vise à proposer des points de

¹ La suite du texte utilise les termes de dispositif médical et technologie de santé, mais les éléments de ce guide sont applicables à toute technologie entrant dans le périmètre d'évaluation de la CNEDiMITS.

repère pratiques sur les aspects méthodologiques pour optimiser le niveau de preuve de ces études et renforcer la confiance dans leurs résultats. Le guide propose un point sur les méthodes actuelles (avantages, inconvénients).

2. Objectif

Ce guide méthodologique, complétant les documents relatifs au [parcours du dispositif médical en France](#) (1) et aux [principes d'évaluation de la CNEDiMTS](#) (2), a pour objectif de répertorier les méthodes (types d'études cliniques et types d'analyses) lorsque la randomisation et/ou l'aveugle sont impossibles à mettre en œuvre en :

- précisant leurs limites ;
- proposant des exemples d'application dans la mesure du possible.

Ce document doit être considéré comme une aide méthodologique et n'est pas opposable. L'objectif n'est pas de proposer une « méthode toute prête » pouvant être appliquée quelle que soit la nature de la demande. La diversité des technologies évaluées, de leur contexte d'utilisation et des populations visées ne permet pas la définition d'une approche unique d'évaluation. Il revient au promoteur de choisir la méthode de son étude clinique en fonction de la nature de la demande, de la nature du dispositif, de la population cible ainsi que de l'arsenal disponible et de la place potentielle de la technologie dans la stratégie thérapeutique de la pathologie concernée.

3. Contexte

3.1. L'étude contrôlée randomisée en double aveugle

Même si les conditions du schéma expérimental peuvent s'éloigner de la pratique clinique courante, les études contrôlées randomisées en double aveugle restent la référence incontournable de l'évaluation de tout produit de santé. En effet, seules la randomisation et la comparaison en double aveugle permettent de rendre les deux groupes étudiés comparables, avec comme seule différence la stratégie allouée, et donc d'attribuer aux stratégies étudiées (produits de santé, interventions...) les différences observées dans les deux groupes en termes de résultats cliniques, dans un schéma thérapeutique donné. Autrement dit, cette méthode permet formellement l'imputabilité des résultats à la technologie évaluée. Le double aveugle permet d'affranchir l'évaluation des biais liés à la subjectivité du suivi, à l'évaluation des critères de jugement, *etc.*

Un outil d'évaluation des risques de biais a été créé pour évaluer le potentiel de biais dans les essais randomisés : l'outil RoB 2.0 fournit un cadre de référence permettant de considérer le risque de biais dans les résultats de n'importe quel type d'essai randomisé. L'outil RoB 2.0 est destiné à être appliqué à des essais randomisés individuels, à des essais parallèles randomisés, à des essais en grappes randomisés et à des essais croisés randomisés (3).

Les cinq domaines suivants sont utilisés dans l'évaluation du risque de biais des essais randomisés individuels :

- les biais qui émergent du processus de randomisation ;
- les biais dus à des déviations par rapport aux interventions prévues ;
- les biais dus à des données finales manquantes ;
- les biais dus à la mesure du résultat ;
- les biais dans la sélection des résultats rapportés.

Le choix d'une méthodologie différente de celle de l'essai contrôlé randomisé doit toujours être argumenté et justifié par l'industriel (comme dans des situations telles que les développements concomitants, des populations particulières pour lesquelles une extrapolation de l'efficacité peut être réalisée sur la base de données observationnelles, *etc.*).

Les méthodes d'évaluation clinique présentées dans ce document sont à envisager lorsque les effets respectifs de deux produits, ou d'un dispositif médical (ou technologie de santé) et d'un placebo, doivent être comparés. 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.*).

3.2. Phases du développement clinique

Le manque de données de bonne qualité constitue un obstacle important dans l'évaluation des dispositifs médicaux et des technologies de santé par les agences d'évaluation et les décideurs (4), (5).

L'évaluation clinique de l'efficacité d'un nouveau dispositif médical, qui est l'objet de ce document, est réalisée après la phase préclinique et les 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 préclinique. 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 à 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 (6) :

- 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, mais également le plateau technique et 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 pour la démonstration du bénéfice.

- La mesure de l'efficacité

L'étude de faisabilité est un préalable pour construire l'hypothèse à tester et calculer le nombre de sujets nécessaires dans une étude comparative. À ce stade du développement, la détermination du critère de jugement principal est essentielle/fondamentale car c'est ce critère qui permettra de mesurer l'efficacité du DM. Le **critère doit être clinique et pertinent** au regard de la pathologie et de l'action recherchée, comme la réduction de la mortalité ou d'une complication cliniquement mesurable. Si un **critère intermédiaire ou de substitution est utilisé, celui-ci devra avoir été validé.**

- Les complications et les risques

À la différence des études pharmacologiques n'impliquant pas d'acte médical, deux types d'évènements indésirables peuvent être rapportés lorsqu'il s'agit de DM : 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.

4. Particularités méthodologiques pour le développement clinique des dispositifs médicaux

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 détaillées ci-dessous (7-10).

4.1. 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 important de son développement clinique. Il est préférable de privilégier l'évaluation clinique le plus précocement possible, en tenant compte dans le protocole d'étude des évolutions possibles de la technologie dès lors qu'elles ne modifient pas sa fonction principale. De plus, le nouveau règlement européen 2017/745 relatif aux dispositifs médicaux impose un renforcement des prérequis nécessaires à l'obtention du marquage CE et des exigences sur le niveau de démonstration du bénéfice/risque. Enfin, lorsqu'un DM est largement diffusé, il est difficile de faire adhérer les professionnels de santé à un protocole d'étude, car une technique déjà utilisée est souvent empiriquement considérée comme efficace (11).

4.2. Population éligible et recrutement

La faible taille de la population éligible peut constituer également une spécificité des études portant sur les dispositifs médicaux. En effet, la population concernée peut être restreinte, une technologie pouvant ne concerner que quelques centaines de patients dans certains cas (12). Dans ce cadre, un essai classique en groupes parallèles peut être plus difficile à mettre en place en raison de sa complexité pour recruter les patients et de son coût.

Le choix de la population étudiée est important (13). 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 de l'étude 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 dispositif médical. Le choix des critères d'inclusion et de non-inclusion dans l'étude est un des points clés du protocole de l'étude.

4.3. Acceptabilité

L'acceptabilité de l'étude pour les patients est déterminante pour sa faisabilité.

L'obtention du consentement du patient est le prérequis pour la réalisation d'essais cliniques. L'étape d'information des patients avant l'obtention de leur 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 professionnels de santé 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 le domaine de la pédiatrie.

Certains patients peuvent avoir une préférence pour une intervention et refusent la randomisation. Des questions d'acceptabilité de l'étude par le professionnel de santé peuvent également se poser lorsque l'utilisateur professionnel est intimement persuadé que la technique qu'il a l'habitude de pratiquer constitue la meilleure stratégie (14). Ces considérations peuvent gêner le recrutement des patients et rendre la randomisation difficile.

4.4. Randomisation

Les techniques de randomisation jugées adéquates sont l'utilisation de tables de nombres aléatoires, la génération informatisée du groupe de randomisation par exemple (15). 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. **Quelle que soit la méthode de randomisation, ses modalités doivent être décrites dans le protocole.**

Bien que capitale, la méthode de randomisation est seulement décrite et adéquate dans à peine la moitié des essais non médicamenteux et l'assignation dans moins d'un quart des cas (16, 17). Plusieurs facteurs pourraient expliquer l'absence de randomisation dans certains essais. D'une part, un essai randomisé est plus coûteux qu'une série de cas (18). D'autre part, la randomisation peut être jugée *a priori* impossible, le plus souvent pour des raisons pratiques (préférence du patient ou du professionnel pour un nouveau dispositif médical potentiellement efficace...). Plus rarement, la comparaison avec un traitement invasif a conduit certains auteurs à ne pas randomiser, mais cet argument est très discutable (19). En effet, la littérature fournit de nombreux exemples montrant que de tels essais sont effectivement possibles dans la plupart des cas, y compris pour évaluer des traitements invasifs *versus* non invasifs, comme ceux destinés à la revascularisation par angioplastie ou par pontage coronaire *versus* traitement médical (20, 24). La comparaison avec un traitement invasif est par ailleurs préférable à une diffusion de dispositifs médicaux sans preuve de leur efficacité et sans évaluation robuste du rapport bénéfice/risque (4, 21).

4.5. 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 tous les intervenants dans la chaîne de soins ou une partie seulement : le patient, le médecin administrant le traitement, la personne recueillant le critère de jugement ou le statisticien. 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.

Boutron *et al.* ont synthétisé les différentes méthodes d'aveugle utilisées dans les essais non pharmacologiques publiés dans les revues avec un *impact factor* élevé (22). L'aveugle peut être complet, partiel ou ne porter que sur l'évaluation du critère de jugement.

L'aveugle est plus souvent impossible dans les essais non pharmacologiques (23), pour des raisons éthiques ou pratiques (4). Dans ces situations, des alternatives ont été développées afin d'évaluer, de la manière la plus objective possible, l'efficacité d'un traitement non pharmacologique.

Dans certains cas, l'aveugle de ces personnes clés est impossible. L'absence d'aveugle peut conduire à des biais tels les biais de sélection, de suivi, d'attrition ou de mesure, remettant en cause la validité interne d'une étude (24, 25).

Des revues des études ayant soutenu l'approbation FDA d'un dispositif médical à haut risque ont montré que : dans le domaine orthopédique entre 2001 et 2015, 76 % des études étaient randomisées, 60 % étaient en aveugle dont 62 % en double aveugle (26), dans le domaine otolaryngologique, entre 2000 et 2014, 46 % des études étaient randomisées et 43 % étaient en aveugle (27). Soixante-dix pour cent (70 %) des études concernant les dispositifs médicaux thérapeutiques soumises au comité d'éthique de Berlin entre 2010 et 2013 étaient randomisées et 38 % en aveugle (16). Une autre revue de la littérature a pu montrer que l'aveugle concernait à peine un quart des patients, 6 % des médecins et deux tiers des évaluateurs dans les essais non pharmacologiques (17). Ceci est particulièrement dommageable, d'autant plus que l'influence du médecin était également bien plus marquée que dans les essais pharmacologiques (17). Or, les essais en ouvert surestiment de 14 % la taille de l'effet par rapport aux essais réalisés en double aveugle (28).

Tableau 1. Exemples d'impossibilité de réalisation de l'étude en aveugle, d'après (29)

Aveugle du patient	Aveugle de l'opérateur	Aveugle de l'équipe assurant le suivi	Évaluation en aveugle
<ul style="list-style-type: none"> – Aspect ou perception du dispositif – Cicatrice qui révèle le type d'intervention (ex. : chirurgie <i>versus</i> traitement non interventionnel, laparotomie <i>versus</i> cœlioscopie...) 	<ul style="list-style-type: none"> – Intervention chirurgicale – Aspect ou manipulation du dispositif 	<ul style="list-style-type: none"> – Effets indésirables spécifiques et caractéristiques d'un des traitements administrés – Coloration ou traces caractéristiques laissées par le dispositif – Aspect en imagerie échographique ou radiologique évocateur d'un des traitements reçus – Cicatrice qui révèle le type d'intervention 	<ul style="list-style-type: none"> – Organisation pratique : nombre excessif de visites ou de consultations, limitant l'acceptabilité pour le patient – Problème d'acceptabilité pour le chirurgien, et également pour le patient, si la consultation est uniquement réalisée par une personne extérieure

4.6. Choix du groupe de référence ou comparateur

Le choix du groupe de référence ou comparateur est crucial et un **comparateur actif** est à privilégier chaque fois qu'il existe (30).

Cette problématique ne concerne pas les essais dans lesquels le dispositif médical évalué s'ajoute au traitement de référence (essais « *on top* » ou « *add on* »). Pour valider la nouvelle technologie que l'on

veut ajouter à la stratégie thérapeutique (ou diagnostique ou de compensation du handicap), le schéma placebo **on top**, stratégie habituelle, est utilisé lorsqu'il n'existe aucune alternative.

L'usage du placebo (ou d'un traitement inactif) peut également être considéré acceptable dans les circonstances suivantes :

- en l'absence de traitement ou d'intervention efficace ;
- lorsque l'abstention d'un traitement ou d'une intervention 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 intervention efficace ne fournit pas de résultat scientifiquement fiable et que l'administration d'un placebo n'ajoute pas de risque important de dommage irréversible (30).

Une revue de la littérature a synthétisé les **principaux placebos de chirurgie** utilisés dans les essais non pharmacologiques (22). 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 ; dans d'autres cas, 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 peut être utilisée dans les cas où il n'y a aucun comparateur convenable, *i.e.* un traitement de référence actif, et lorsqu'elle comporte peu de risques (24, 31-37). À noter, la standardisation des soins préopératoires (même position des patients ou du matériel), peropératoires (durée de l'intervention, instruments, manipulation ou soins) et postopératoires est également importante. Dans d'autres études, le chirurgien ayant fait le geste ne participe pas au suivi des patients. Cela peut être une solution en l'absence de placebo possible. Boutron *et al.* ont également rapporté les différents placebos possibles pour l'utilisation des dispositifs médicaux : prothèses placebos, dispositifs médicaux identiques mais inactivés, ou activés mais rendus inefficaces, ou utilisation simulée d'un DM (22, 38).

Tableau 2. Exemples de placebos utilisés en chirurgie et pour les dispositifs médicaux, d'après (22) et (29)

Placebo		
Techniques chirurgicales	En période anesthésique	Patients sous anesthésie générale Patients masqués par un drap
	Pendant l'intervention	Incisions cutanées pour obtenir une cicatrice similaire à celle de l'intervention testée (exemple : pontage aortocoronarien avec une artère mammaire interne <i>versus</i> incisions cutanées seules chez les patients avec une angine de poitrine, chirurgie arthroscopique <i>versus</i> incisions cutanées seules chez des patients atteints d'arthrose) Simulation de l'intervention (exemple : injection intracérébrale de cellules fœtales <i>versus</i> incisions cutanées et abrasion de la corticale externe de la boîte crânienne chez des patients atteints de la maladie de Parkinson)
	En postopératoire	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 cœlioscopique, ou pansements similaires pour comparer exérèse du côlon par laparotomie et colectomie cœlioscopique Standardisation des soins postopératoires/traitements associés
Dispositifs	DM placebo	Placebo non strictement identique au traitement reçu

médicaux	<p>Prothèse similaire mais ne fournissant pas l'effet thérapeutique (par exemple sans production de chaleur)</p> <p>Utilisation de DM différents cachés (exemple : sac de transfert pour aphérèse témoin, lunettes opaques pour le participant...)</p>
En postopératoire	<p>Utilisation d'un appareil identique inactivé (même durée, fréquence, précautions...)</p> <p>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</p>

4.7. Facteurs liés à l'expérience des opérateurs

4.7.1. 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 (4). 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) (39).

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 (24) qu'il est nécessaire de prendre en compte, par exemple par une période d'entraînement (40). Lors des interventions chirurgicales, l'impact de l'apprentissage peut être évalué par des indicateurs (comme le taux de réussite de mise en place du dispositif, le taux de succès technique de l'intervention...).

La variabilité de l'effet du dispositif médical selon le niveau d'expertise de l'opérateur doit également être évaluée. Cette variabilité est fréquente et de nombreux exemples pourraient l'illustrer, comme dans les essais comparant angioplastie et thrombolyse dans le traitement de l'infarctus du myocarde (41, 42). Aussi, l'évaluation d'une nouvelle technologie *versus* la technique de référence risque d'être déséquilibrée en défaveur de la nouvelle technique, en raison de l'expérience de l'opérateur. Une évaluation trop précoce risque de refléter les complications liées à l'apprentissage de la nouvelle intervention. L'évaluation doit donc intégrer l'effet de l'apprentissage, par exemple en enregistrant la formation et l'expérience (courbe d'apprentissage) (24). Ainsi, 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.

4.7.2. Volume des actes

Après avoir acquis une expertise technique initiale, l'entraînement régulier de l'opérateur peut avoir un impact. Le volume d'activité annuel de l'opérateur (et du centre) doit donc 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 professionnel de santé (43). Cette association est d'autant plus importante pour les interventions chirurgicales complexes, rarement effectuées, et à risque (44).

4.8. Facteurs d'ajustement

Dans un essai comparatif *a fortiori* randomisé, il n'existe pas de problème en termes de confusion car les autres composantes sont identiques dans les deux groupes (toutes choses égales par ailleurs). En effet, l'utilisation de la technologie de santé doit être standardisée car elle devra être impérativement reproduite dans l'usage futur du DM. Mais il arrive parfois qu'il soit difficile d'isoler les effets des technologies de santé de ceux des autres composants du système de soins. En effet, les DM sont souvent utilisés en association avec d'autres interventions (chirurgie, diagnostic, surveillance...). L'évaluation de l'effet propre du DM sur le critère de jugement peut donc être difficile. Les interventions chirurgicales dépendent non seulement du chirurgien (préférence, apprentissage et volume d'activité...), mais également des pratiques pouvant varier dans les différents centres, des autres membres de l'équipe (anesthésiste, infirmier), de la prise en charge pré et postopératoire et de l'organisation des soins. Le déploiement de technologies connectées introduisant une profonde refonte de l'organisation des soins, voire du parcours de soins, questionne de fait sur ce qui doit être évalué : le DM seul ou le DM dans une organisation de soins (4), (24)-(11).

Il est important de définir *a priori* très précisément dans le protocole ce qui doit être évalué et, le cas échéant, l'encadrement nécessaire à l'utilisation du DM ainsi que l'organisation des soins 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 (13). Cependant, dans le cas où ces pratiques ne pourraient être reproductibles, on cherchera à recueillir les éventuelles différences entre les centres plutôt qu'une standardisation de l'organisation d'emblée, afin de pouvoir ajuster l'analyse sur ces éléments.

4.9. Types d'analyse

4.9.1. Analyse en intention de traiter

L'analyse en intention de traiter consiste à analyser les patients dans leur groupe initial de randomisation, quel que soit le traitement qu'ils ont finalement effectivement reçu et quelle que soit leur évolution par rapport à l'étude. Elle est à privilégier dans les essais de supériorité car il s'agit de l'attitude la plus conservatrice. Elle permet de ne pas perdre le bénéfice de la randomisation et de limiter les effets qui ne seraient pas dus au hasard mais seraient en rapport avec l'intervention.

4.9.2. Analyse en per-protocole

L'analyse en per-protocole consiste à analyser les patients selon le dispositif médical qu'ils ont effectivement reçu, et non selon ce qui était initialement prévu pour leur groupe initial de randomisation. Cette stratégie peut entraîner une augmentation de la différence de l'effet des stratégies comparées. Elle est à privilégier dans les essais de non-infériorité.

5. Études expérimentales

5.1. 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 dispositif médical 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 dispositif médical 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 dispositif médical reçu.

Dans les interventions chirurgicales, l'aveugle du patient peut être réalisé si le patient est sous anesthésie générale ou si le geste est masqué par un champ opératoire (22).

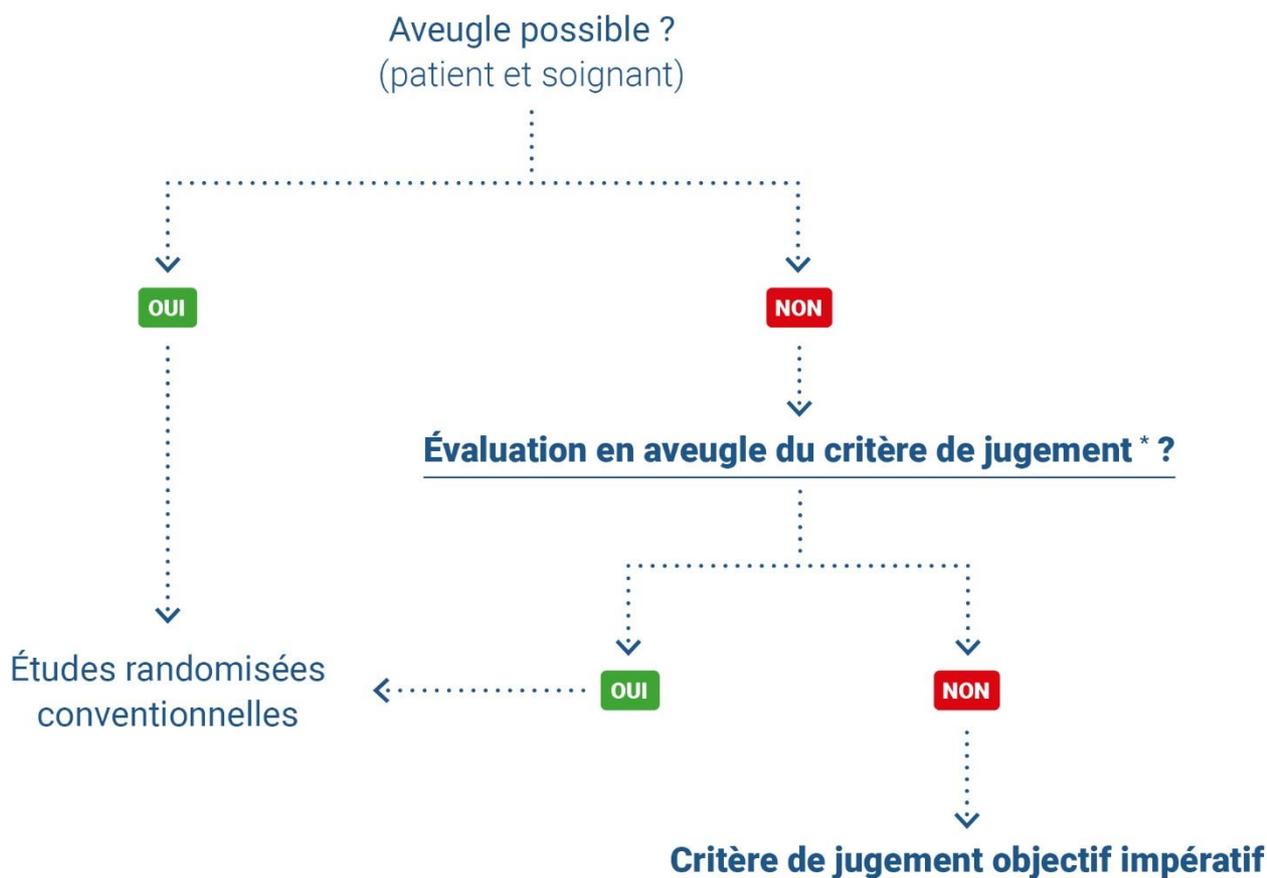
Dans certains cas, un comité d'adjudication indépendant des investigateurs est mis en place afin de contrôler le critère de jugement. L'aveugle des hypothèses de l'étude (ou aveugle partiel) est également une alternative décrite (22). Elle repose sur une information partielle délivrée aux participants, 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 motifs 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'experts indépendants).

La figure 1 présente l'arbre décisionnel afin de garantir 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

5.2. 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 durant 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 dispositif médical 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 dispositif médical et non pour le traitement de référence (45). Ce type d'étude est **en pratique très peu utilisé** à cause des contraintes méthodologiques comme les biais de sélection, mais également de problèmes éthiques spécifiques du fait de l'absence d'information donnée dans le groupe randomisé pour recevoir le traitement conventionnel.

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 dispositif médical devront signer le consentement éclairé. En cas de refus, le patient recevra le traitement de référence (46).

Trois possibilités ont été proposées par Zelen.

- (1) En cas d'éligibilité, le patient est randomisé dans l'un des deux groupes suivants : le groupe G1 dans lequel le consentement n'est pas demandé et dans lequel les patients reçoivent le traitement de référence et sont informés sur le traitement ; et le groupe G2 dans lequel le consentement est demandé et dans lequel les patients reçoivent le dispositif médical expérimental (figure 2)(45).
- (2) 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) (figure 3)(45).
- (3) 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. Cette alternative est particulièrement intéressante lorsqu'il n'y a pas de traitement de référence établi « double consentement » (figure 4) (47).

Figure 2. Prérandomisation d'après Zelen, cas 1 (45)

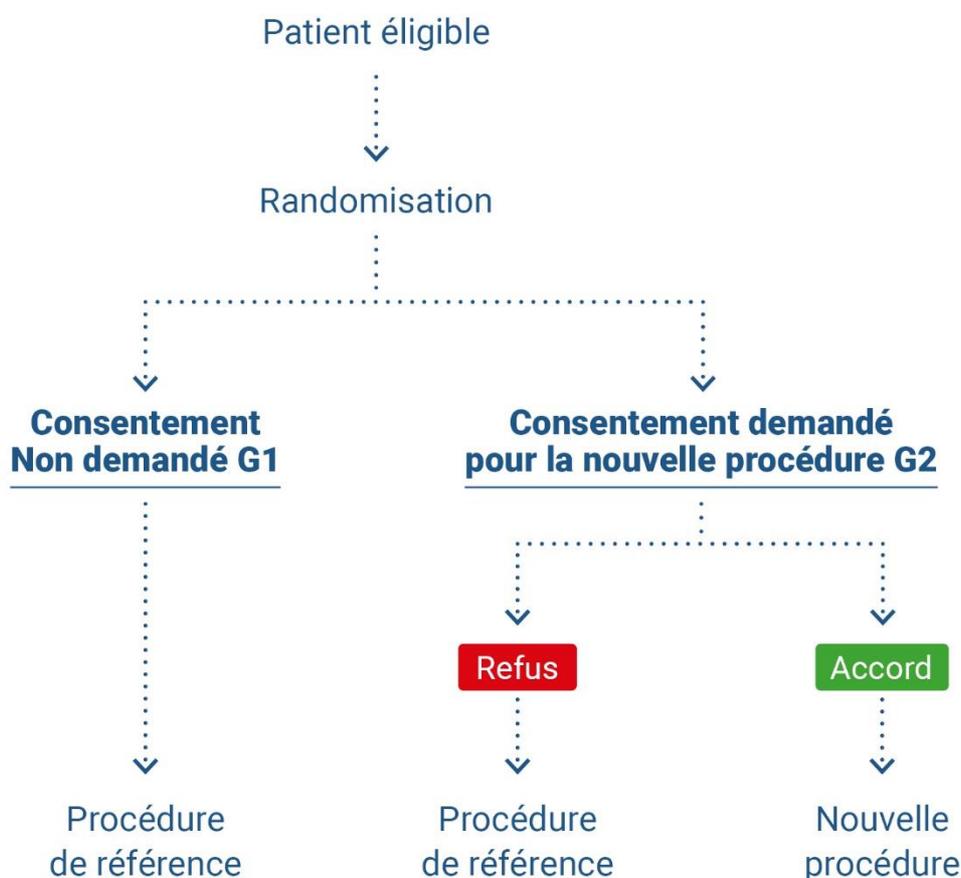


Figure 3. Prérandomisation d'après Zelen, cas 2 (45)

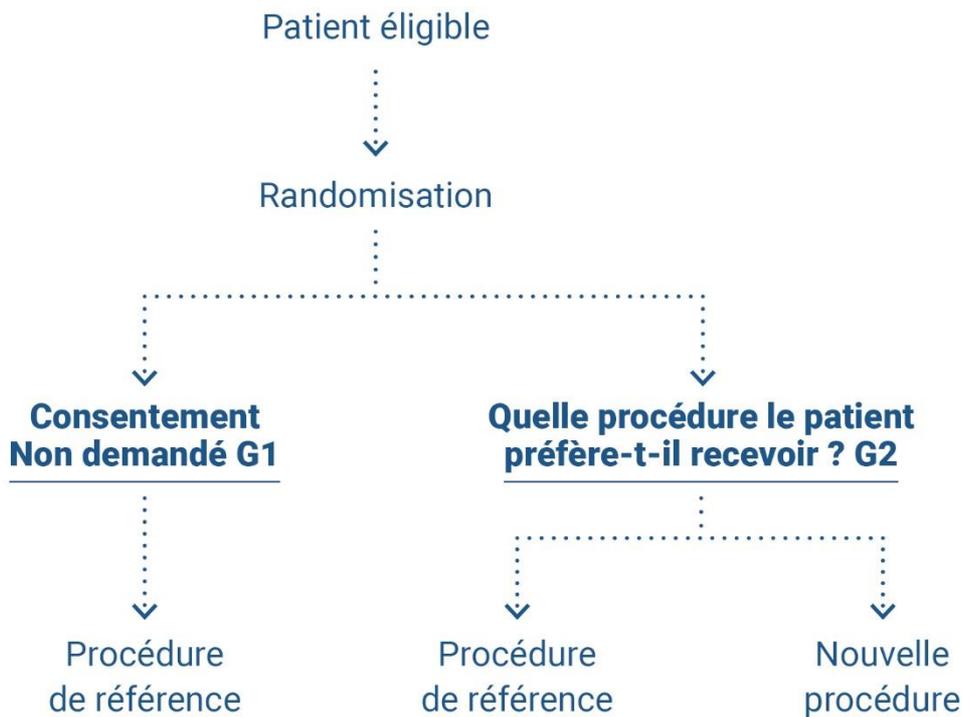
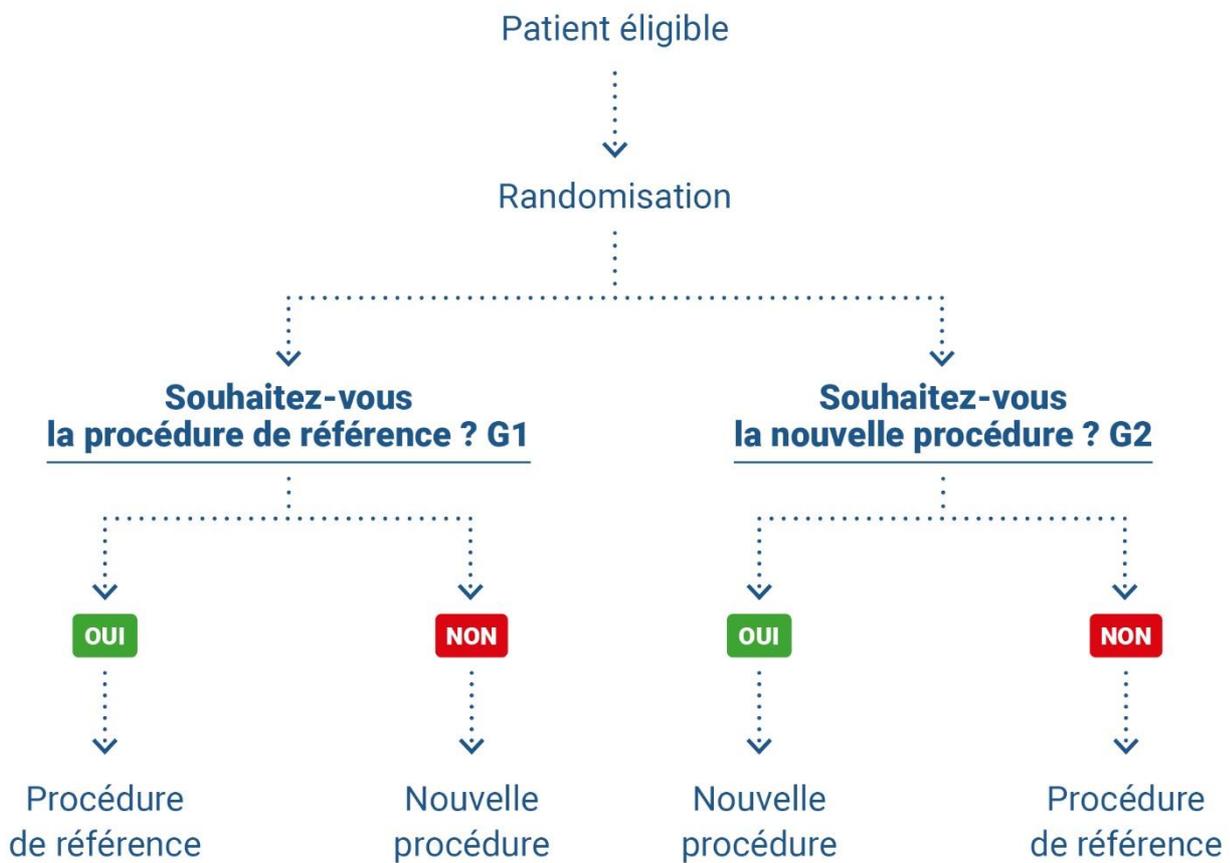


Figure 4. Double consentement d'après Zelen, cas 3 (47)



Avantages

Le patient connaît le traitement qui lui sera administré avant de donner son consentement, ce qui évite le retrait de consentement une fois le traitement attribué. Ce type d'étude exige une décision du patient uniquement lorsque celui-ci doit recevoir le dispositif médical expérimental, ce qui permet de limiter le stress du patient. 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, dans la mesure où le médecin présente un seul traitement au patient. Ce plan d'étude peut ainsi faciliter l'inclusion des patients et, finalement, permettre d'inclure plus de patients que dans un essai conventionnel (45), (48).

Inconvénients

Un essai randomisé avec prérandomisation 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 (49). Du fait de l'absence de consentement éclairé dans le groupe G1, ce plan ne peut pas être utilisé dans les essais contre placebo. 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 habituelles. 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 (50).

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 dispositif médical peut être masquée si beaucoup de patients du groupe G2 choisissent le traitement de référence (50). Ainsi, le fait de comparer le groupe G1 au groupe G2 quel que soit le traitement reçu dilue l'effet mesurable du nouveau dispositif médical (45). Ainsi, il existe un risque de ne pas pouvoir conclure.

Enfin, le double aveugle est impossible (45) 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'atteinte². 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 G1 avec celles des patients du groupe G2 recevant le traitement de référence à la suite de leur refus de recevoir le nouveau dispositif médical, afin d'identifier des biais de sélection potentiels (45).

² Or, il n'y a pas de sous-représentation dans le groupe témoin puisque le consentement n'est pas demandé dans ce groupe.

Champs d'application

Très peu utilisé en France, ce plan expérimental peut être 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 ; ainsi que dans les domaines de la chirurgie, des études interventionnelles ou des dispositifs médicaux ;
- pour des pathologies dans lesquelles la décision du traitement est psychologiquement difficile pour le patient (cancérologie, soins palliatifs, choix du traitement pour son enfant...) (50, 51).

5.3. Essais fondés sur l'expertise (*expertise-based randomised controlled trials*)

Contexte

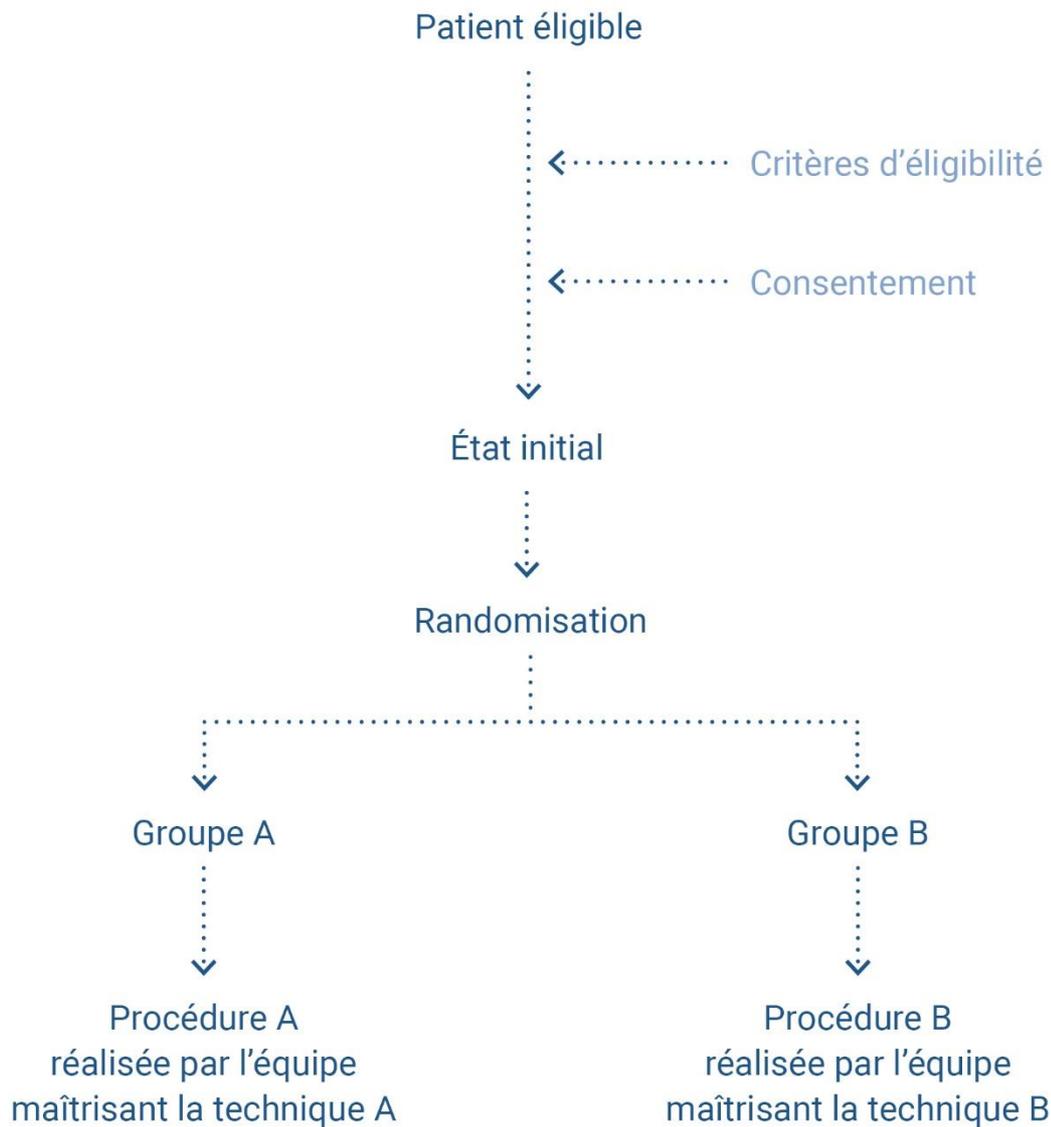
Les essais contrôlés randomisés classiques peuvent être difficiles à mettre en place en particulier dans le domaine interventionnel du fait de problèmes d'acceptabilité pour l'opérateur. Une mauvaise acceptabilité peut gêner le recrutement des patients dans un essai clinique conventionnel (14). De plus, l'opérateur pourrait inconsciemment réaliser de manière plus méticuleuse la procédure chez les patients randomisés pour l'intervention qu'il préfère ou réaliser un suivi différent que des professionnels de santé moins expérimentés pour la technique évaluée (14, 52). 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 fondés sur l'expertise (*expertise-based randomised 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 fondé sur l'expertise consiste à randomiser les patients auprès du chirurgien ou de l'équipe qui maîtrisent l'intervention dédiée (figure 5) (14). Le chirurgien ou l'équipe ne réalisent qu'une seule procédure, dont ils ont la maîtrise et pour laquelle ils sont experts. Cela suppose que la maîtrise de l'intervention a été préalablement vérifiée et qu'elle est comparable dans chaque groupe (39). Ce type d'essai peut limiter l'acceptabilité pour le patient qui doit accorder sa confiance à deux chirurgiens ou à deux équipes.

Figure 5. Schéma représentant la randomisation dans les essais contrôlés randomisés basés sur l'expertise, d'après (39)



Avantages

- Amélioration de la validité interne

Les essais contrôlés randomisés fondés sur l'expertise offrent une meilleure validité interne (52) pour les raisons suivantes.

- Absence de biais d'expertise différentielle

Les essais contrôlés randomisés fondés sur l'expertise permettent de réduire le biais d'expertise différentielle qui existe dans les essais conventionnels (53).

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

Ce type d'essai permettrait également de limiter les biais de suivi différentiel 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 serait plus limité que dans un essai conventionnel dans lequel les chirurgiens pourraient suivre de manière différente les

patients des différents groupes (52). 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 œuvre de l'essai.

- Réduction des écarts au protocole

Afin de pouvoir évaluer au mieux l'efficacité d'un nouveau dispositif médical, 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.* (52) ont montré dans un essai clinique conventionnel sur le traitement des fractures tibiales (alésage vs non-alésage) des écarts au protocole différents 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 lorsque les chirurgiens pratiquent peu l'intervention dans laquelle le patient a été randomisé que dans le cas contraire (54-56). De ce fait, le risque d'écart différentiel au protocole est réduit dans les essais basés sur l'expertise, le chirurgien réalisant uniquement l'intervention qu'il a l'habitude de pratiquer (52).

- Amélioration de la faisabilité

La faisabilité d'un essai randomisé fondé sur l'expertise est meilleure que celle des essais conventionnels. En effet, les chirurgiens n'ont pas besoin d'avoir été formés à deux interventions et seront plus facilement recrutés avant de débiter l'essai clinique (52).

- Meilleure acceptabilité par les chirurgiens

Une enquête transversale a montré que 58 % des chirurgiens orthopédiques préféreraient participer à des essais contrôlés fondés sur l'expertise, contre 17 % pour les essais cliniques contrôlés randomisés conventionnels. La préférence était influencée par l'expertise des chirurgiens (14). 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 et pour laquelle il est « expert » (39, 52). Au total, les chirurgiens sont moins réticents à participer à un essai clinique fondé sur l'expertise plutôt qu'à un essai conventionnel (39).

Inconvénients

L'aveugle n'est pas possible et le risque de biais de mesure est majeur (52). 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 (14). 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 (24, 53).

Contraintes

Ce type de schéma nécessite au moins la présence d'un expert de chacune des interventions randomisées dans chaque centre (14). De plus, la consultation initiale de pré-inclusion doit être réalisée par une personne neutre qui déterminera l'éligibilité des patients (14).

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 (52). 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 fondés sur l'expertise sont potentiellement intéressants dans toutes les situations où il y a un intérêt avéré à 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 – ou lorsque la maîtrise de la technique a une influence importante sur le résultat – et lorsqu'elles sont réalisées par des types de professionnels de santé différents (57, 58).

Cette méthode pourrait également être utilisée pour des techniques déjà diffusées pour lesquelles la randomisation n'est plus possible compte tenu de la conviction des utilisateurs, ou lorsque les pratiques sont différentes selon les centres.

5.4. *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 (59). 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 (24).

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 (24).

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 le principe d'équilibre, également appelé « équilibre » (absence d'arguments pour conseiller un traitement plutôt qu'un autre), l'utilisation maximale des différentes données disponibles. Ce type d'essai permet d'évaluer rapidement des techniques nouvelles ou innovantes ou potentiellement dangereuses (59).

Inconvénients

L'analyse est complexe du fait de la prise en compte de l'expérience de l'opérateur et de 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 (59).

L'organisation pratique est difficile (24, 59), d'autant plus quand la technologie évolue, car la plupart des opérateurs sont alors peu expérimentés (59).

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 (59).

Champs d'application

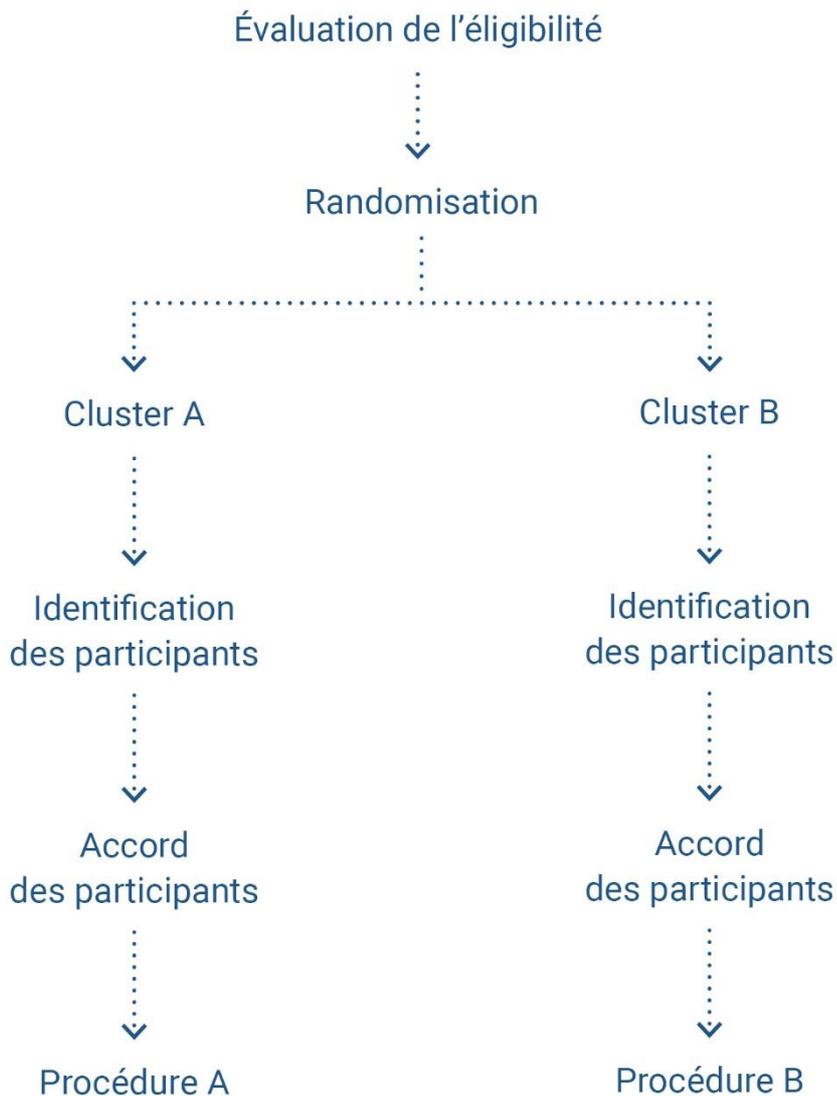
Actuellement, il existe peu d'exemples disponibles dans la littérature sur ce type d'étude. Il semblerait présenter plutôt un intérêt en phase préliminaire, pour écarter des technologies dangereuses, et donc en amont de l'évaluation de l'efficacité clinique objet de ce guide.

5.5. 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) (60). L'unité de randomisation est le centre de soin et non le patient. Les patients inclus dans l'essai recevront le traitement pour lequel le centre qui s'occupe d'eux a été randomisé. 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 (61)



Deux types de plan d'essai en *cluster* ont été décrits : les essais en *cluster* complet ou les essais en *cluster* avec recrutement actif (62). Dans les essais en *cluster* complet, une personne dans chaque centre a la charge du *cluster* et se voit attribuer la procédure à tester.

Avantages

Ce design permet de minimiser le biais de contamination quand le traitement en aveugle n'est pas possible. La randomisation de *clusters* plutôt que d'individus permet de prévenir une éventuelle contamination entre deux techniques disponibles dans un même centre (62-64). Ainsi, lors de l'utilisation d'un dispositif médical en postopératoire, le risque d'administrer le dispositif A au lieu du dispositif 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 (62, 64).

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, l'absence de connaissance sur le *cluster* dans lequel chaque patient aurait dû être inclus rend l'analyse en intention de traiter difficile à mettre en œuvre 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 (62). Si dans les essais à unité de randomisation individuelle ce risque est faible car le nombre d'unités de randomisation est élevé (fréquemment supérieur à 100 ou 200), ce risque est majeur dans les essais à unité de randomisation collective où l'on randomise parfois moins de 10 unités (hôpitaux, structures de soins ou médecins). Bien évidemment, plus le nombre d'unités de randomisation disponibles est faible, plus le risque d'un déséquilibre dû au hasard entre les groupes témoins et les groupes intervention est élevé (64).

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 (62). 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 (62). Aucune méthode statistique ne permet de limiter ce biais. De plus, l'influence de l'unité de randomisation (individu ou groupe d'individus) est beaucoup plus forte dans un essai en *cluster* que dans un essai conventionnel (62).

Pour les essais en *cluster* avec recrutement actif, il peut également y avoir des responsables du *cluster* inactifs 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 prédéterminé, ce qui peut induire un biais de sélection et un recrutement différentiel (62). 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 (62).

Les essais en *cluster* soulèvent également des questions éthiques, puisqu'ils supposent deux niveaux de consentement : celui des professionnels de santé, puis l'information des patients sur le fait qu'ils participent à une étude (65). Ainsi, les patients n'auront pas le choix du traitement. Dans les essais où un nouveau dispositif médical 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 compromet ainsi l'aveugle du patient (65). En l'absence d'aveugle, il n'est pas rare dans les études en *cluster* que les hôpitaux, services ou médecins qui se voient assigner une intervention qui ne leur convient pas (par exemple s'ils sont randomisés dans le groupe témoin) refusent finalement de participer à l'étude, aggravant le risque de déséquilibre entre les groupes (64).

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 (62). L'analyse statistique doit également tenir compte de cet effet *cluster* (62).

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 (62).

Afin de limiter les biais de sélection et le recrutement différentiel dans les essais en *cluster* avec recrutement actif, Puffer *et al.* (66) 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 œuvre pour des raisons logistiques. Des 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 (62).

Champs d'application

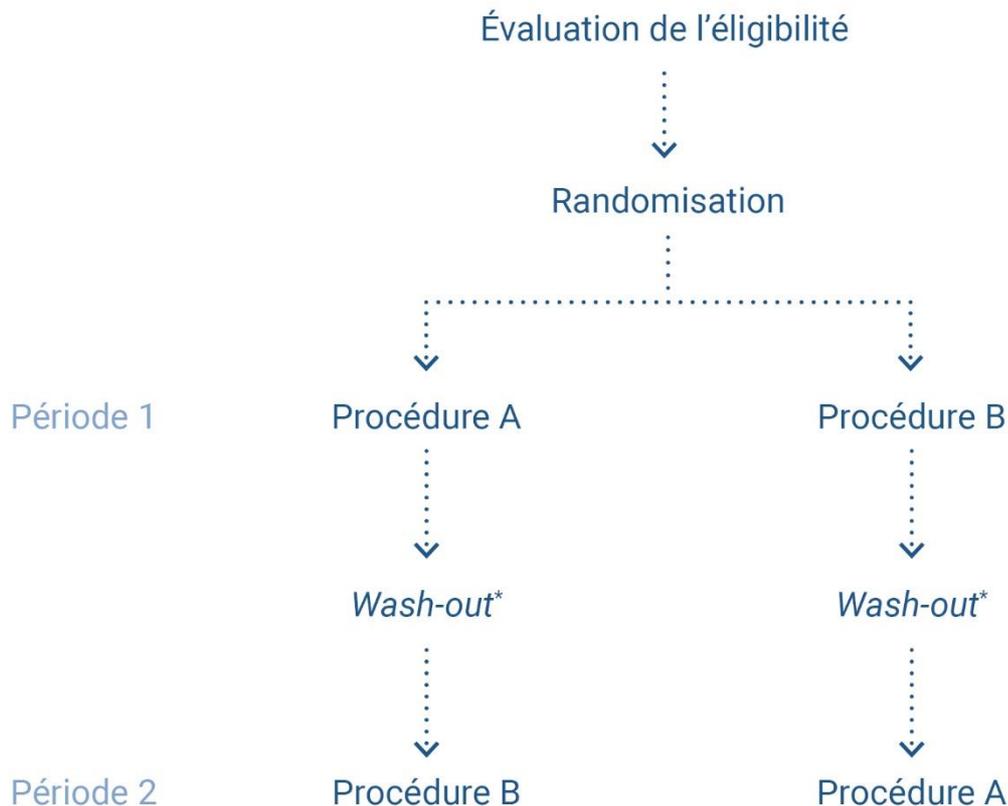
Les essais en *cluster* sont essentiellement utilisés pour les interventions organisationnelles, comportementales ou encore les programmes de promotion pour la santé (64). Dans la mesure où ils sont susceptibles d'améliorer la faisabilité ou les aspects pratiques, ils pourraient être envisagés dans le 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 (51, 67).

5.6. 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) (68). Chaque patient reçoit les deux traitements (69), qui sont ainsi comparés sur le même patient au cours de différentes périodes (70).

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 (70), 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 (69). 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 premier 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 irréversible (la guérison ou le décès) (69) ou lorsqu'il y a une courbe d'apprentissage. Dans le cas où les mesures effectuées chez les mêmes patients 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.

La durée des traitements comparés est plus longue que dans un essai en groupes parallèles, exposant ainsi à un plus grand risque de perdus de vue. La mesure du critère de jugement peut être modifiée par un effet « d'apprentissage » de certains tests d'évaluation (biais de mesure).

Contraintes

L'ensemble des biais doit être discuté (68). 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. 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 (69).

L'effet des traitements ne doit pas être irréversible afin que les sujets se retrouvent en début de seconde période dans un état identique à celui qu'ils avaient en début de première période.

La maladie ne doit pas évoluer de façon notable entre les deux périodes. Par exemple, elle ne doit pas guérir (ou s'aggraver) spontanément avant la fin de la deuxième période.

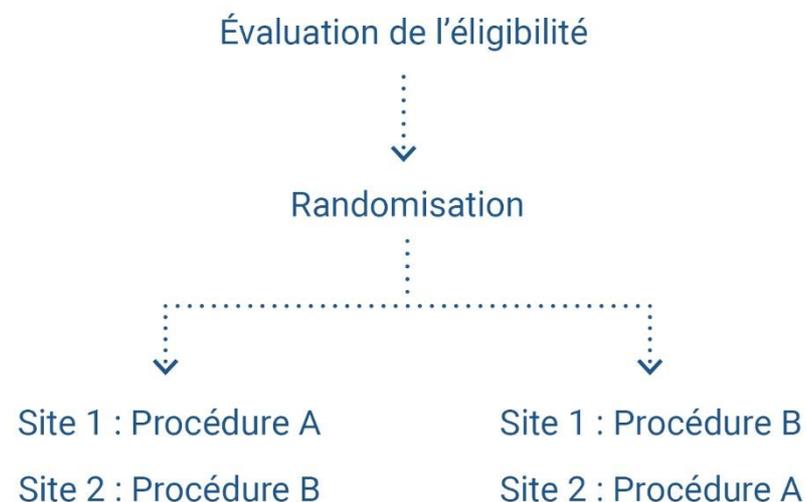
Il ne doit pas y avoir d'interférence entre l'ordre d'administration des traitements et leur effet. L'effet d'un traitement doit être le même que celui-ci soit administré en premier ou en second. Pour vérifier ce point, il est nécessaire de rechercher une interaction entre le traitement et la période (69) ou effet *carry-over*. Si celui-ci est présent, seules les données issues de la première période seront analysées.

Du point de vue statistique, des tests spécifiques pour séries appariées doivent être utilisés pour tenir compte du fait que chaque sujet est son propre témoin.

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* peut être utilisé dans l'évaluation de dispositifs médicaux destinés à la prise en charge de maladies chroniques d'évolution stable, dispositifs dont les effets disparaissent rapidement à l'arrêt de leur utilisation et dont l'effet est d'apparition rapide. Il existe également des essais en *cluster* en *cross-over* (71).

5.7. Études expérimentales en cas unique (SCED) [Nouveauté]

Contexte

Certains dispositifs médicaux sont indiqués pour une population cible de **très petite taille** (de quelques dizaines à quelques centaines de patients), rendant difficile le recrutement d'un large échantillon de patients lors des études visant à démontrer leur efficacité. Les études expérimentales en cas unique ou SCED (*Single Case Experimental Design*) sont un ensemble de schémas d'études qui permettent de s'adapter à cette difficulté de recrutement. Elles visent aussi à pallier les biais méthodologiques liés au profil hétérogène des patients et à l'absence de groupe contrôle.

Ce type d'étude, utilisé depuis une cinquantaine d'années dans le domaine de l'éducation et de la psychologie, est actuellement exploré dans le domaine de la médecine physique et de réadaptation (MPR), que ce soit pour démontrer l'intérêt d'interventions de rééducation ou de certains dispositifs médicaux, en pratique clinique pour déterminer la prise en charge adaptée à un patient donné ou dans le cadre d'études expérimentales à des fins de généralisation des résultats.

Principe

La méthodologie SCED repose sur trois principes (72) :

- l'étude intensive et prospective d'une seule personne ou d'un petit groupe de sujets au cours du temps, chaque patient étant son propre contrôle ;
- la mesure répétée et fréquente du critère de jugement durant toutes les phases de l'étude ;
- l'introduction séquentielle et randomisée de l'intervention au cours de l'étude.

Les études de type SCED comportent au moins deux phases :

- la phase de ligne de base ou *baseline* (désignée par la lettre A) durant laquelle les performances de chaque patient font l'objet de plusieurs mesures. Cette étape permet d'identifier une ligne de tendance ;
- la phase de traitement ou d'intervention (désignée par la lettre B) durant laquelle les mesures répétées continuent.

L'introduction de l'intervention se fait de façon séquentielle et randomisée pour chaque sujet. L'entrée séquentielle des patients dans le protocole vise à contrôler l'absence d'effet *retest* ou de récupération spontanée en ligne de base.

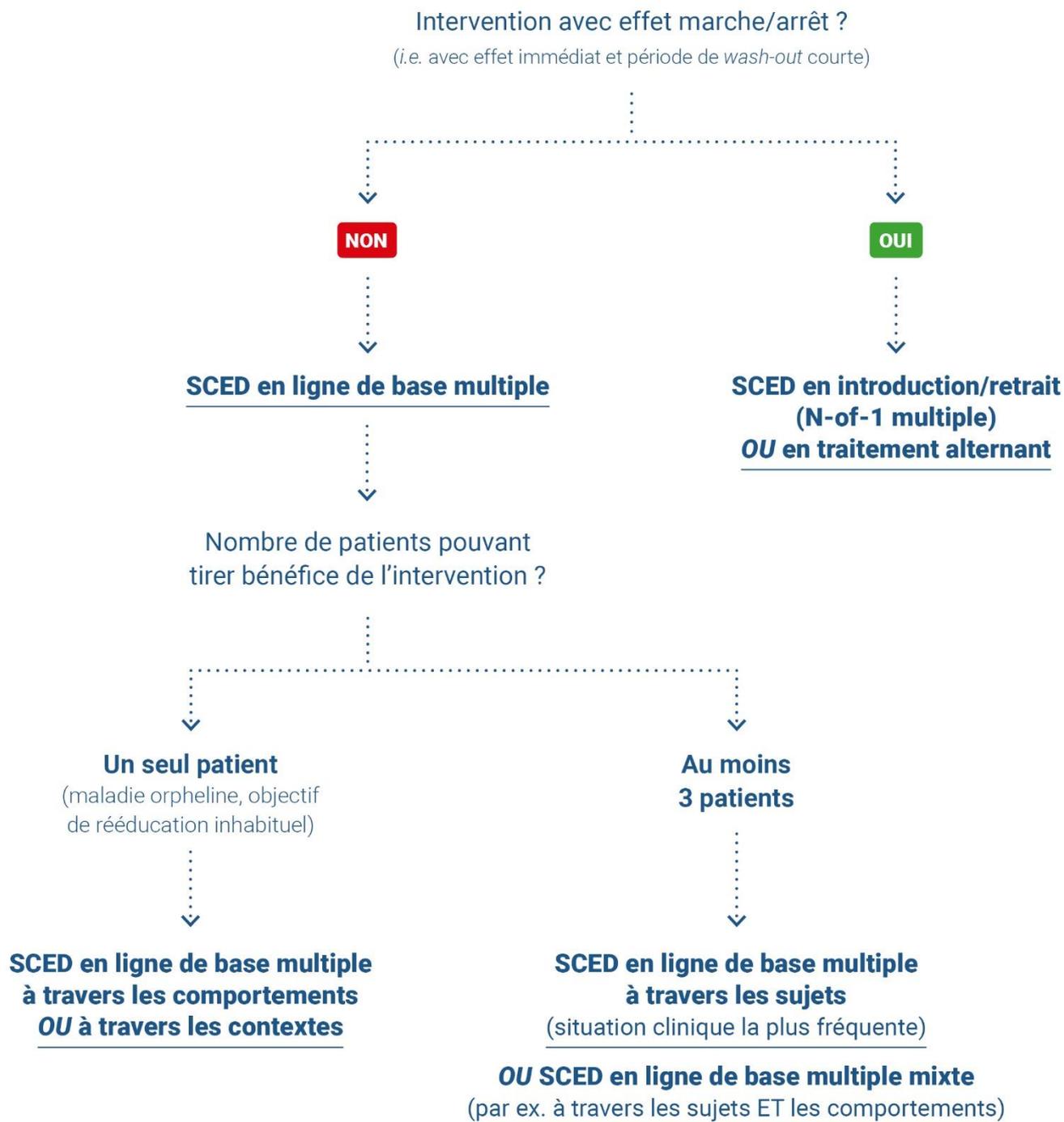
Le schéma expérimental peut inclure plus de deux phases avec une alternance de phases de lignes de base et de phases d'intervention (schéma de type ABAB). Il n'est pas nécessaire que les interventions soient identiques.

Trois principaux types de plan d'essai SCED ont été décrits (72) :

- les essais en introduction/retrait ABAB, aussi appelés essais de type N-of-1 multiple, pour les cas où l'effet de l'intervention est immédiat, de type marche/arrêt et non rémanent, avec une période de *wash-out* courte ;
- les essais SCED en traitement alternant, pour les cas où l'effet de l'intervention est immédiat et de type marche/arrêt avec une période de *wash-out* courte et lorsque l'on souhaite étudier au moins deux interventions différentes ;
- les essais SCED en ligne de base multiple : à travers les sujets (au moins 3 sujets), à travers différents contextes (par ex. d'abord en centre de rééducation, puis à la maison,

puis à l'école) ou à travers les comportements (introduction séquentielle de plusieurs modules d'une intervention ciblant différents comportements). Ce type de SCED s'applique quand l'effet de l'intervention n'est pas immédiat et est durable, et que le changement induit par l'intervention est *a priori* lent.

Figure 9. Choix du type de SCED, d'après (72)



La mesure du critère de jugement est répétée au moins trois fois (souvent plus) avant d'introduire l'intervention, puis répétée pendant l'intervention, juste après et à distance, le même nombre de fois à chaque phase de l'étude. Pour pouvoir conclure sur l'efficacité de l'intervention, le plan d'étude doit prévoir au moins trois démonstrations de l'effet (réplications) à trois différents points dans le temps. Cette recommandation de « trois démonstrations » ne repose pas sur une base formelle, mais il s'agit d'une norme conceptuelle retrouvée dans la littérature traitant de la méthodologie des études de type SCED (73).

Avantages

La réplication de la démonstration de l'effet et la randomisation du moment de l'introduction de l'intervention et de l'ordre des patients contribuent à la bonne validité interne de ce type d'étude. Certains auteurs recommandent au moins cinq mesures par phase pour renforcer la validité interne et la validité statistique du résultat (73, 74).

Chaque sujet étant son propre contrôle, ce type de *design* permet de ne pas avoir recours à l'appariement des sujets. Dans une étude de type SCED, la puissance vient du nombre de mesures répétées et non du nombre de patients inclus.

Inconvénients

La question de la généralisation des résultats obtenus avec ce type d'essai est un point essentiel qui doit être considéré spécifiquement lors de la conception de l'étude. Afin de renforcer la validité externe d'une étude de type SCED, il est recommandé de répliquer l'étude : au terme des premiers patients, le même protocole doit être appliqué de nouveau, à l'identique, sur le même nombre d'autres patients, et ceci au moins trois fois, dans plusieurs centres différents (73, 74).

Contraintes

La principale contrainte est la nécessité de définir un critère de jugement approprié pour évaluer l'effet de l'intervention. La méthodologie SCED exige généralement que les cliniciens créent un critère de jugement *ad hoc* pertinent et qui peut être mesuré de manière fiable de nombreuses fois.

Le choix de la durée des phases est également un élément important pour la réussite de l'étude.

Différents outils ont été développés pour la conception et l'analyse critique de ce type d'étude :

- les recommandations du *What Works Clearinghouse* pour la documentation technique des SCED (73) ;
- un guide pratique spécifique aux SCED destinés à évaluer des interventions en réadaptation (72) ;
- l'échelle RoBiNT (*Risk of Bias in N-of-1 Trial scale*) pour évaluer la qualité méthodologique d'une étude de type SCED (75).

Sous réserve de respecter ces règles de conception et de réalisation d'étude (fidélité procédurale) et même si les exemples d'applications dans un contexte d'évaluation des technologies de santé sont encore limités, les études de type SCED représentent actuellement une perspective d'études de bonne qualité méthodologique avec peu de patients.

Champs d'application

Les études de type SCED pourraient notamment être utiles pour évaluer l'efficacité d'aides techniques, d'orthèses, de prothèses externes dans le champ de la médecine physique et de réadaptation, ainsi que pour l'évaluation d'autres types de DM destinés à compenser des déficiences tels que les aides auditives.

5.8. Études *in silico* [Nouveauté]

Contexte

Les études *in silico* sont définies comme « l'utilisation de la simulation informatique individualisée pour le développement ou l'évaluation réglementaire de produits, de dispositifs médicaux ou d'interventions médicales » (11, 76)³. Il s'agit donc de modélisations utilisées en tant qu'**études de faisabilité** pour guider les stratégies de développement. **Ils ne remplacent pas les essais cliniques pour la démonstration de l'efficacité du produit de santé chez l'homme.**

Principe

Les études *in silico* regroupent un très large ensemble de méthodes numériques qui, utilisant des modèles mathématiques, permettent de simuler l'effet d'un produit de santé sur une pathologie et sur une population de patients virtuels à l'aide de l'outil informatique. Elles nécessitent des données obtenues d'après des études *in vivo* ou *in vitro*, ou d'après des observations d'études cliniques, et les relient par des corrélations statistiques aux informations structurales. La qualité de ces études dépend donc de la qualité de la base de données utilisée (77, 78).

Avantages

Les études de faisabilité *in silico* ont un intérêt pour aider à définir la stratégie de développement des technologies de santé lorsqu'il existe un modèle physiopathologique, car elles sont peu coûteuses et rapides à réaliser.

Inconvénients

Malgré les efforts mis en œuvre pour recueillir des données homogènes, la validité de ces études ne peut être vérifiée, notamment parce que ces études *in silico* reposent pour la plupart sur des études préliminaires avec de forts risques de biais.

Contraintes

Actuellement aucune étude *in silico* n'a permis de remplacer l'intégralité du développement clinique et l'évaluation d'un produit de santé sur un organisme entier. Un modèle physiologique de la pathologie est parfois nécessaire. Une combinaison des méthodes *in silico*, *in vitro* et *in vivo* reste nécessaire. Ces essais ne sont pas destinés à remplacer les essais cliniques chez l'homme, mais permettent d'optimiser les essais cliniques conventionnels pour l'évaluation clinique des dispositifs médicaux⁴ (79).

Les études *in silico* sont actuellement réservées aux études de faisabilité et jouent un rôle d'orientation, très tôt dans le processus de développement des produits de santé. Cette approche vise à optimiser les études traditionnelles *in vitro* et *in vivo*. Par exemple, elle permet d'explorer des hypothèses par simulation numérique, avant le lancement d'essais cliniques conventionnels, pour augmenter leur

³ How simulation can transform regulatory pathways : <https://www.fda.gov/science-research/about-science-research-fda/how-simulation-can-transform-regulatory-pathways>

⁴ Computational modeling and simulation (CM&S): paving the way for evaluation of 21st century medical devices. Medical Device Innovation Consortium <https://mdic.org/program/computational-modeling-and-simulation-cms/>

probabilité de succès ou identifier plus précocement des échecs. Elles permettent d'affiner la stratégie de développement clinique et de répondre à différentes questions selon le stade de développement : preuve de concept, optimisation des études cliniques, analyse décisionnelle lors du développement, par exemple identification de la meilleure indication, de la ou des sous-populations cibles dans des pathologies pour lesquelles il est difficile de recruter des patients, en priorité dans les maladies rares ou pédiatriques (80, 81).

Champs d'application

Les études *in silico* ne peuvent pas être utilisées comme des études d'évaluation car il est impossible de connaître leur validité. Ce sont des travaux d'aide à la construction des études. Dans ce cadre, le développement et la validation de modèles *in silico* sont encouragés par la FDA et l'EMA, afin d'identifier les éventuelles perspectives d'extrapolation de leurs résultats pour l'évaluation clinique des produits de santé.

Les études *in silico* peuvent être envisagées pour les situations suivantes (82, 83) :

- information sur la conception des essais cliniques à mettre en œuvre ;
- identification de la population cible la plus pertinente pour la réalisation d'une étude ;
- prévision de la tolérance du dispositif médical.

Concernant les dispositifs médicaux, ces études sont parfois utilisées pour évaluer la dynamique des fluides pour prédire comment le sang ou d'autres fluides corporels se déplacent à l'intérieur et autour du dispositif testé, ou pour analyser des éléments structurels afin de confirmer que les forces échangées entre le corps et le dispositif médical ne causeront aucun dommage.

À ce jour, ces essais font partie de l'arsenal disponible pour le développement clinique des produits de santé, en tant qu'outils complémentaires aux outils existants permettant d'optimiser la phase de développement clinique d'un produit de santé lorsqu'il existe un modèle physiopathologique. Elles sont actuellement réservées uniquement aux études de faisabilité.

5.9. Types d'analyse

5.9.1. Méthodes bayésiennes [Nouveauté]

Contexte

L'inférence bayésienne est une alternative à l'approche fréquentiste pour l'analyse des études (randomisées ou observationnelles). Les méthodes d'analyse bayésiennes peuvent être utilisées dans toutes les indications et phases et elles ne sont pas limitées aux petits effectifs ou aux essais séquentiels (84).

Dans les essais randomisés conventionnels de type fréquentiste, les données apportées par des études antérieures sont utilisées uniquement au moment de la conception de l'étude (c.-à-d. ne sont pas prises directement en compte dans l'inférence). Par la suite, seule l'information collectée au cours de l'essai est utilisée. Les résultats antérieurs peuvent également être pris en compte de manière formelle lors de la réalisation d'une méta-analyse.

L'approche bayésienne prend en compte les informations existantes (quelle que soit leur nature) et les informations collectées par l'essai en cours et permet surtout de donner les probabilités (85). Le point majeur à surveiller est le calibrage de la **distribution a priori** qui doit être non informative, accompagnée d'une étude de sensibilité.

Dès 2010 dans les recommandations de la FDA, l'utilisation de l'approche bayésienne dans le domaine des dispositifs médicaux a été considérée comme faisant partie des méthodes possibles, notamment lorsque le nombre potentiel de participants aux essais est limité (86).

Principe

L'approche bayésienne permet de calculer une distribution de la probabilité d'efficacité du traitement à partir des données de l'étude et d'une distribution *a priori*. La distribution *a priori* peut être informative et conditionner le résultat final. Une distribution *a priori* informative permet en théorie de prendre en compte des connaissances préalables sur l'effet du dispositif, mais fait courir le risque d'influencer fortement et arbitrairement le résultat de l'étude. Pour cette raison, il est préférable et il est attendu qu'une distribution non informative soit utilisée.

Elle consiste à utiliser ces informations existantes (cela n'a pas forcément de lien chronologique), qui peuvent être fournies par la littérature, des cohortes ou d'autres essais cliniques. Lorsque cette information est utilisée, une étude de sensibilité, intégrant impérativement un *a priori* réellement non informatif et un *a priori* pessimiste, est nécessaire afin d'évaluer l'influence du choix de la distribution *a priori* sur les résultats potentiels de l'essai.

À noter que, par la structure de l'inférence bayésienne, la distribution *a priori*, combinée à la vraisemblance (les données de l'essai clinique faisant partie de l'expérimentation), donne la distribution *a posteriori* sur laquelle sont tirées les conclusions.

Dans un essai clinique bayésien, l'incertitude sur un paramètre d'intérêt est décrite selon une distribution de probabilités, actualisées au cours de la collecte d'information dans l'essai. Les probabilités constituant la distribution *a priori* sont fondées sur l'information disponible à l'extérieur de l'essai (données de la littérature, des opinions d'experts, des cohortes ou d'autres essais cliniques). Sur la base du théorème de Bayes, les probabilités *a posteriori* sont estimées à partir de l'essai en cours, conditionnellement aux probabilités *a priori*.

L'estimation bayésienne ne fournit pas d'intervalles de confiance, mais des intervalles de crédibilité fondés sur la distribution *a posteriori* et une probabilité *a posteriori* d'efficacité du traitement (85). À la différence de l'approche fréquentiste, il n'y a pas de tests statistiques, mais des résultats s'appuyant sur la distribution *a posteriori* et ses intervalles de crédibilité. Il est aussi possible de calculer des distributions de probabilités prédictives du paramètre d'intérêt afin de conclure l'essai.

Avantages

L'intérêt d'un DM doit être apporté par une étude en elle-même. Les méthodes bayésiennes permettent de compléter l'information en fournissant la probabilité que le dispositif médical puisse être efficace conditionnellement aux connaissances extérieures et aux observations de l'essai. Elles permettent donc 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, cette méthode permet de réduire le nombre de sujets nécessaires, soit grâce à l'apport des informations extérieures, 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. Elle permet aussi d'extrapoler les performances d'une version du

DM à une autre. Enfin, en utilisant des informations extérieures à l'essai, elle peut réduire les effectifs et, ainsi, réduire la durée et les ressources nécessaires d'un essai clinique (85, 87).

Inconvénients

L'inconvénient majeur est le risque de mauvais calibrage de la distribution *a priori*, influençant ainsi le résultat final de l'essai. Ceci peut être évité par la réalisation d'une étude de sensibilité sur le choix de l'*a priori* avant son utilisation.

Contraintes

Les analyses bayésiennes reposent sur la construction de distribution(s) de probabilités *a priori* à partir d'informations de nature et de qualités potentiellement variables. Il est donc important de faire appel à des méthodologistes qui ont l'expertise dans ce type d'analyse. 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 de plus en plus**, en cardiologie notamment (88-94).

5.9.2. Essais séquentiels

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.

Contexte

Les essais séquentiels ont été développés afin de permettre de ne pas donner un traitement moins efficace à des patients plus longtemps que nécessaire (95).

Principe

Les différentes étapes sont planifiées indépendamment des données collectées auparavant (96).

Différentes méthodes d'analyses séquentielles ont été décrites :

- la méthode séquentielle classique reposant sur l'analyse de paires de patients (peu utilisée actuellement) (97) ;
- les analyses séquentielles en groupe reposant sur l'analyse tous les n sujets (également peu utilisée de nos jours) (98-101) ;

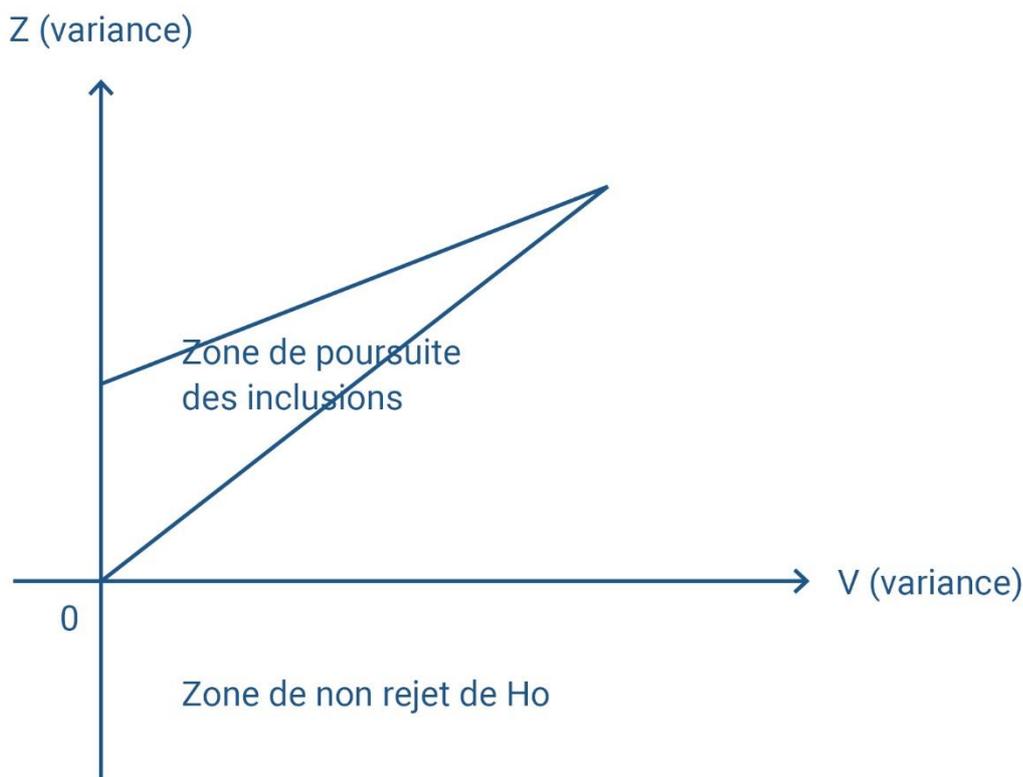
- les méthodes flexibles permettant de choisir à quel moment faire les analyses (57-58, 79, 80, (102) ;
- et les analyses séquentielles continues comme le test séquentiel de rapport de probabilité et le test triangulaire (103-105), (106), (105, 107).

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 (108).

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 bêta. 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 (95). Dans ces deux cas, les inclusions sont stoppées.

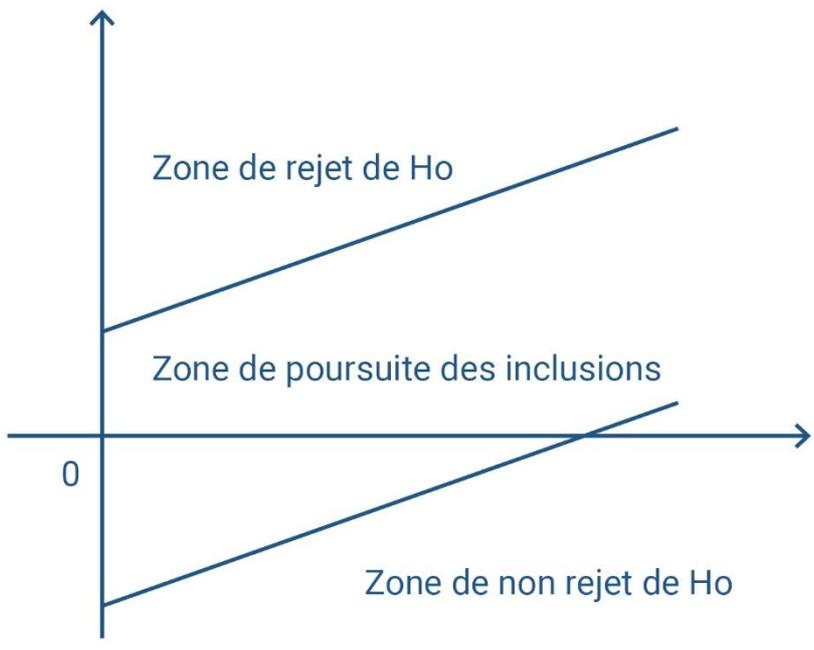
Les figures ci-dessous représentent les règles de décision graphique pour arrêter ou pour poursuivre les inclusions.

Figure 10. Règles de décision graphique pour le test triangulaire (plan fermé), d'après (107, 109)



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

Figure 11. Règle de décision graphique pour le test séquentiel de rapport des probabilités (plan ouvert)



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 (108). En outre, le nombre moyen de sujets nécessaires dans ce type d'essai est plus faible que dans les essais conventionnels, indépendamment de la taille de l'effet ou de la puissance (95, 110).

Inconvénients des essais séquentiels

Les méthodes séquentielles ont des contraintes pratiques. 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 dispositif médical sera évalué (108). 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 (95). 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 (95).

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 (95). 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 (95).

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) (95). 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 prendre en compte/mesurer d'éventuels effets indésirables.

L'utilisation de méthodes spécifiques pour le calcul des intervalles de confiance (103, 104) 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.

5.9.3. Essais adaptatifs

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 (108).

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 (111).

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 (95) sans augmentation du risque alpha (112). Toute l'information collectée lors des différentes étapes est utilisée, et pas seulement celle de l'étape en cours (96). L'allocation du traitement d'un patient dépend des résultats obtenus avec les patients précédents (112).

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 (95).

Les essais adaptatifs peuvent être en deux phases ou en plusieurs phases (113). Les essais adaptatifs en deux 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 deux phases.

Avantages

Les essais adaptatifs permettent une grande flexibilité. De plus, toute l'information collectée est utilisée pour l'analyse (95, 96). 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 (108, 114).

Inconvénients

Un essai adaptatif requiert des contraintes logistiques plus grandes qu'un autre type d'essai (96). Il exige également une grande vigilance concernant la réestimation 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 (108). Comme pour les essais séquentiels, il est nécessaire de définir un critère de jugement simple et univoque de l'efficacité (108). 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 (95).

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 (97, 108).

Du fait de ces inconvénients, les essais adaptatifs ont été jugés peu intéressants par rapport aux essais séquentiels (108, 115, 116).

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) (95, 117).

Enfin, l'enregistrement prospectif des hypothèses et des adaptations réalisées au fur et à mesure est capital pour la crédibilité des conclusions finales (95).

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.

6. Études en vie réelle ou études observationnelles [Nouveauté]

6.1. Principes généraux [Nouveauté]

Contexte

Le terme méthodologique consacré à ce type d'étude est « **études observationnelles** », car on ne fait, en théorie, « qu'observer » une population ou un phénomène sans intervenir, ou en intervenant le moins possible, sur leur évolution naturelle. On désigne également ce type de données, qui ne sont pas collectées dans un cadre expérimental mais générées à l'occasion des soins réalisés en routine sans intervention sur les modalités usuelles de prise en charge des patients, par le terme « **données de vie réelle** » qui reflètent donc *a priori* la pratique courante.

Les « **études en vie réelle** » ou études observationnelles ont un intérêt évident en tant qu'études complémentaires d'études expérimentales pour recueillir des données en situation réelle d'utilisation. En effet, la force de l'essai clinique constitue également sa faiblesse. À titre illustratif, la rigueur du schéma expérimental qui est nécessaire à la démonstration de l'efficacité peut entraîner des conditions parfois éloignées de la pratique clinique et remettre en cause la transposabilité des résultats à la pratique courante. Les résultats obtenus sur des durées courtes peuvent être peu compatibles avec l'évolution rapide des technologies et des stratégies.

La prise en compte des données collectées en conditions réelles d'utilisation est donc un enjeu majeur dans l'évaluation des produits de santé. En améliorant leur recueil, stockage, analyse et transparence, et plus globalement la confiance que l'on peut porter à leur résultat, la pertinence de l'évaluation des produits de santé au service des patients ne sera que renforcée (118). La HAS a développé un guide spécifiquement dédié aux « Études en vie réelle pour l'évaluation des médicaments et des dispositifs médicaux » (119, 120-121)

Principe

Les études en vie réelle peuvent répondre à des enjeux très différents (évaluation clinique, épidémiologique, évaluation du fardeau d'une maladie, etc.). Elles peuvent permettre de comparer des groupes de sujets traités aux sujets non traités (ou traités par une alternative), afin de mettre en évidence l'association entre le traitement considéré (dispositif) et l'évolution de la pathologie. Elles peuvent permettre la mesure de l'impact d'un dispositif sur la morbi-mortalité ou la qualité de vie dans des conditions moins contrôlées que celles des essais cliniques classiques interventionnels (122).

Ces données en vie réelle, anonymisées, peuvent être issues de plusieurs sources : d'études observationnelles spécifiques créées pour la circonstance fondées sur un recueil *ad hoc*, d'études à partir de bases de données existantes, des registres, des données issues de dispositifs connectés ou collectées sur le Web, ou des combinaisons de ces modes de recueil (comme les cohortes chaînées avec les données des bases médico-administratives), etc. (121).

Dans une situation de comparaison, il est absolument indispensable d'explicitier l'essai que l'on vise à mimer (« **The Target Trial** ») *i.e.* l'étude contrôlée randomisée hypothétique qu'il serait souhaitable de mener afin de pouvoir répondre à l'unique objectif de causalité identifié (« **émulation de l'essai cible** ») (123-127).

Les standards méthodologiques actuels des études en vie réelle doivent prendre en compte les éléments suivants :

- un seul et unique objectif principal doit être défini par étude ;
- les critères d'inclusion et de non-inclusion doivent être identifiés avec attention ; et il est préférable d'inclure des patients nouveaux utilisateurs du dispositif ;
- il est préférable de choisir un comparateur actif (la décision de traiter exclut le temps avant le traitement) ;
- le choix des variables à prendre en compte est crucial (le choix des variables associées au pronostic du critère de jugement est plus important que les variables différenciant les groupes comparés) ; et les variables doivent être de qualité (standardisation par exemple) ;
- la vérification de l'absence de biais de confusion résiduel doit être systématique ;
- le temps de début de suivi (T0 ou J0) doit être défini correctement (risque de biais de temps immortel notamment par comparaison à un placebo (« pas de traitement spécifique »)) ;
- l'analyse doit préférentiellement être réalisée en intention de traiter (ITT) ;
- le plan d'analyse statistique (SAP) doit être préétabli avant l'analyse.

Avantages

En général, les études en vie réelle peuvent inclure des patients qui n'auraient pas pu l'être dans des essais randomisés car ne satisfaisant pas à tous les critères d'inclusion (en particulier dans le groupe non traité avec la technologie de santé, ce qui n'est pas profitable dans le cas des méthodes d'analyse mesurant un *treated average/average treatment effect* comme le *matching*). Elles peuvent permettre d'obtenir des données relatives à des populations de petites tailles, qui n'auraient pas pu être incluses ou qui auraient été sous-représentées dans les essais cliniques randomisés. Elles permettent un suivi plus long. Elles peuvent révéler des complications qui n'auraient pas pu être observées dans les essais cliniques. Elles renforcent la validité externe des résultats (119, 120).

Si les conditions sont comparables et les méthodologies bien conçues, essais cliniques et études en vie réelle peuvent aboutir à des résultats similaires. Un travail de la Cochrane Collaboration en 2014 conclut que les résultats retrouvés dans les études en vie réelle bien conduites peuvent être comparables aux résultats des essais randomisés contrôlés, mais pas toujours. Il existe des exemples pour lesquels les résultats d'études en vie réelle et d'essais randomisés divergent (128-130).

Inconvénients

La qualité des études en vie réelle et la validité scientifique de leurs résultats sont des enjeux cruciaux pour leur prise en compte optimisée dans l'évaluation des produits de santé par la HAS. Les sources de données utilisables sont diverses par leur origine (dossiers médicaux, bases médico-administratives, données d'activités hospitalières/pharmaceutiques, registres, études observationnelles, recensements des effets indésirables, données épidémiologiques, démographiques et environnementales, résultats déclarés par les patients, sondages de santé, réseaux sociaux...) et leur qualité (131).

Contraintes

L'importance des choix méthodologiques à opérer dans les études en vie réelle est déterminante pour optimiser leur niveau de preuve, les approches pouvant être différentes selon la question posée (132, 133). Il n'existe aucune méthode parfaite, chacune présentant des avantages et des inconvénients. Le choix de la méthode doit être discuté et argumenté en fonction du contexte dans lequel s'inscrit l'étude (pathologie et produit de santé étudiés, contraintes administratives et réglementaires, données disponibles dans les bases existantes, etc.). **En pratique tous les biais possibles doivent être discutés lors de la conception du protocole de l'étude en présentant la méthode adoptée pour les contrôler et en cas de contrôle difficile ou impossible, une analyse de sensibilité, prévue au protocole, doit permettre d'établir dans quel sens et dans quelle mesure ces biais ont pu modifier les résultats et les conclusions de l'étude** (122). L'ensemble des biais doit être évalué par l'utilisation d'outils adaptés (par exemple ROBINS-I) (134, 135).

Champs d'application

La HAS publie également un guide dédié aux études en vie réelle afin d'optimiser le niveau de preuve de ces études et la confiance dans leurs résultats et qui détaille leur champ d'application.

Les études en vie réelle concernent principalement les études générées après la mise sur le marché en France et sont plutôt destinées à argumenter des demandes d'extension d'indications ou de renouvellement d'inscription. Pour les innovations en particulier, les premières données étant souvent limitées, le recueil de données de ce type est essentiel pour conjuguer mise à disposition précoce et évaluation (136).

Il est toutefois rappelé que les données en vie réelle sont également prises en compte dans les premières évaluations des produits de santé par la HAS, notamment celles générées dans le cadre du forfait innovation ou de la prise en charge transitoire des dispositifs médicaux. De plus, des données en vie réelle sur l'épidémiologie, le fardeau de la maladie ou celles constituées à des fins de comparaisons indirectes peuvent également contribuer aux évaluations de la HAS.

Figure 12. Pourquoi mettre en place une étude en vie réelle, d'après le guide de la HAS « Études en vie réelle pour l'évaluation des médicaments et des dispositifs médicaux »



6.2. Études observationnelles comparatives spécifiques à une collecte prospective des données

Contexte

Les études de cohorte sont classiquement utilisées à des fins de recherche épidémiologique. Elles peuvent toutefois être utilisées lorsque les essais contrôlés randomisés ne peuvent pas être envisagés pour des raisons éthiques par exemple, lors du cycle de développement d'un dispositif médical (108), dans le cas des études post-inscription en vue d'un renouvellement d'inscription sur la LPPR ou dans le cadre de comparaisons indirectes.

Principe

Les études de cohorte permettent de réaliser des analyses comparatives afin de déterminer si les patients « exposés » au produit de santé, comparativement aux patients « non exposés » (ou exposés aux comparateurs), ont un risque relatif inférieur, égal ou supérieur de présenter l'évènement considéré avec possibilité d'une inférence causale du lien (si le contrôle des biais est approprié, notamment le contrôle des facteurs de confusion).

Les études de cohorte incluent un échantillon de patients aux caractéristiques communes, représentatif de la population visée, qui est suivi longitudinalement afin d'observer la survenue d'évènements de santé dans le temps au sein de cette population définie. Elles comparent l'incidence du critère de jugement entre un groupe recevant le dispositif médical A et un groupe ne recevant pas le dispositif médical A (soit il bénéficie d'une autre intervention, 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 (137, 138).

Figure 13. Exemple d'étude observationnelle prospective comparant la survenue d'infections postopératoires selon le type de suture utilisé



Avantages

Les études de cohorte offrent la possibilité de répondre à des questions que les essais cliniques classiques n'explorent pas, notamment sur les éléments contextuels qui peuvent interagir avec l'effet propre du traitement et en modifier l'efficacité en pratique courante de soins. Elles peuvent permettre de constituer des groupes de sujets exposés de taille suffisante et peuvent également tenir compte de phénomènes liés au temps.

Une réalisation prospective permet de planifier le recueil de toutes les informations nécessaires. Bien conduites, ces études peuvent limiter les biais d'information, lorsque les renseignements sur l'exposition sont enregistrés prospectivement et chez tous les sujets à risque d'exposition. Elles peuvent permettre également de recueillir des informations précises sur l'intervention (par exemple, doses administrées dans un appareil d'imagerie ou durée d'une intervention). De plus, la mesure de l'incidence d'un évènement, d'un risque relatif et la prise en compte du délai de survenue de l'évènement après l'intervention sont possibles.

Inconvénients

Les principaux risques des études en vie réelle sont les biais de sélection qui surviennent quand la probabilité d'inclusion est influencée par la maladie et l'exposition (139-141). En effet, leur principale limite est la non-comparabilité des groupes : les patients exposés et non exposés n'étant pas – par construction – tirés au sort, les caractéristiques différentes des deux groupes peuvent expliquer, au moins en partie, les écarts observés (8, 142). En effet, alors que l'objectif serait d'être comparable sur les variables qui sont susceptibles d'influencer le critère de jugement principal, les limites principales résident essentiellement dans l'érosion de représentativité au fil du temps. En particulier, le biais de déplétion des susceptibles est un biais de sélection lié aux traitements passés des patients éligibles à l'étude. Il peut conduire à surestimer le bénéfice d'un dispositif (ou à en sous-estimer le risque). En effet, les patients ayant pris le traitement d'intérêt par le passé et l'ayant arrêté (par exemple, pour cause d'effet indésirable ou d'un manque d'efficacité) ne seront pas comptabilisés comme « exposés » au traitement en question au début de l'étude. En d'autres termes, la cohorte théorique des patients exposés au traitement d'intérêt s'est « vidée », au fil du temps, des patients pour lesquels le traitement n'était pas adapté. Ce risque de biais rend pertinent le recours aux patients incidents (*new-user*

design), qui consiste à n'inclure que des patients utilisant le DM d'intérêt pour la première fois de leur vie (143, 144). Dans une étude de cohorte, le biais de temps immortel correspond à la prise en compte dans les analyses d'un temps « d'exposition à un traitement » qui inclurait en réalité la période où les patients du groupe « exposé » n'avaient pas encore ce traitement, c'est-à-dire la période entre l'évènement ayant conduit au diagnostic et l'initiation effective du traitement. Une des mesures pour pallier ce biais est de réaliser les analyses selon les « personnes-temps ». Cette période ne doit pas être prise en compte dans le nombre de « personnes-temps » exposées mais ajoutée au nombre de « personnes-temps » non exposées (124).

La faisabilité de ce type d'étude est à discuter selon le niveau de diffusion de la technologie. En effet, en cas de nouvelle technologie, si le produit est très attractif il sera utilisé de manière quasi exclusive chez tous les patients (il n'y aura plus de patients non traités pouvant servir de contrôle). De ce fait, ces études peuvent être lourdes et coûteuses à mettre en place (145, 146). Par ailleurs, pour ces études, le monitoring des patients doit être renforcé. Enfin, étant donné que les données sont collectées au sein d'un établissement de santé et dans un système de santé particulier, et que celui-ci pourrait avoir des pratiques distinctes, il convient d'interpréter les analyses dans le contexte du système de santé dans lequel ces données ont été collectées, et de stratifier les analyses par pays en cas d'étude internationale.

Contraintes

La fiabilité de ces études repose sur une méthodologie rigoureuse. La réalisation du protocole, les analyses et l'interprétation des résultats doivent tenir compte des biais possibles (145-150).

La présence de personnel dédié pour vérifier la qualité des données doit être systématique. 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. Il est donc recommandé de respecter les points suivants (138, 151-155) :

- identifier tous les facteurs de confusion potentiels lors de la planification de l'étude (par un graphique de causalité) ;
- limiter au maximum les patients perdus de vue et les données manquantes en minimisant la pénibilité de participation à l'étude pour les patients ainsi que pour les médecins, à l'aide d'un monitoring adéquat et suffisant des données (avec notamment une recherche active du statut vital des patients) ; et systématiquement prévoir un registre des non-inclusions ;
- sauf demande contraire, s'assurer que l'étude permettra de documenter la population rejointe représentative de tous les patients traités par le produit de santé en pratique courante de soins ;
- lors de l'analyse statistique, comparer, pour étayer la représentativité, les caractéristiques des investigateurs participants aux non participants (les comparer également à l'échelon national), celles des patients inclus aux patients non inclus dans l'étude et comparer les caractéristiques des patients selon leur statut (perdus de vue ou non, données manquantes ou non concernant les principales variables d'intérêt). Suivant les cas, utiliser des techniques d'imputation⁵ ou analyses de sensibilité et utiliser des méthodes de prise en compte des facteurs de confusion (modèle multivarié avec appariement et ajustement sur les facteurs de confusion ou sur les scores de propension par exemple) ;
- procéder à un contrôle de la qualité, un management et des audits des données.

⁵ Avant d'utiliser une méthode d'imputation, il faut vérifier que les données manquantes sont dues au hasard seul.

Champs d'application

Les études observationnelles comparatives ont un intérêt incontestable dans la génération de preuves en dehors du contexte des ECR et sont une source potentielle de connaissance de l'effet des dispositifs médicaux dans des conditions moins contrôlées que celles des essais cliniques classiques (156-160).

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 (avec les limites évoquées plus haut). L'étude de cohorte est par ailleurs essentiellement réservée à l'analyse de maladies assez fréquentes. Les données issues de ces études peuvent être utilisées dans une évaluation à long terme pour un renouvellement d'inscription par exemple. Les études de cohorte historico-prospectives sont inadéquates pour évaluer de nouvelles indications (9, 151, 161).

6.3. Cohortes complètes selon la préférence patient

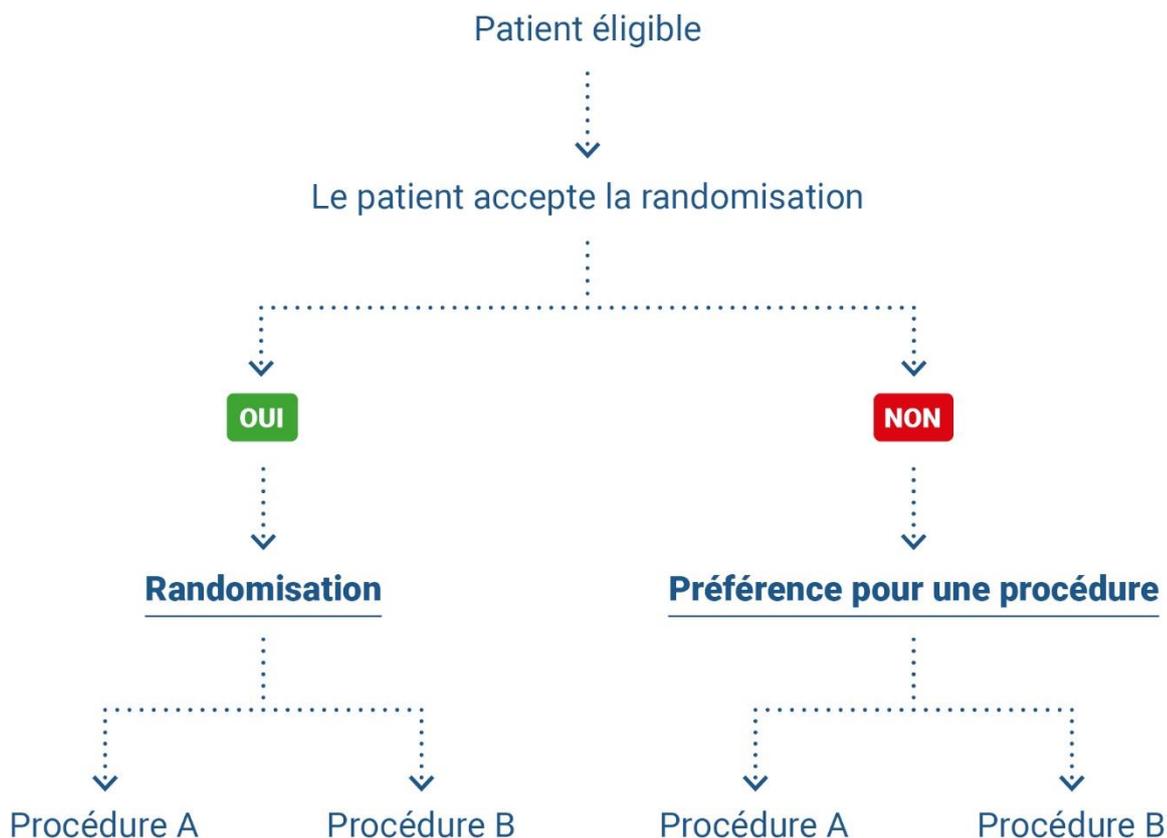
Contexte

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 (162). 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*) basées sur les préférences patients ont été développées 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 (162). Lorsque le patient refuse la randomisation, il reçoit le traitement qu'il préfère (figure 11).

Figure 14. Principe des études basées sur les préférences ou des *comprehensive cohort studies*, d'après (163)



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 (164).

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.

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 peut être particulièrement intéressant en chirurgie, dans les procédures interventionnelles et pour les dispositifs médicaux.

6.4. Études à partir de données existantes [Nouveauté]

Contexte

Les bases de données déjà existantes (« données secondaires »), comme les bases de données médico-administratives, peuvent être utilisées dans les études avec un très bon rapport coût/qualité puisque les données sont déjà enregistrées.

Les principales sources de données secondaires :

- les données médico-administratives ou autres bases de données intégrées dans le *Health Data Hub* en France ;
- les données issues de la production de soins en routine :
 - o les systèmes d'information hospitaliers, en particulier les entrepôts hospitaliers de données,
 - o les systèmes d'information publics ou privés en médecine de ville, notamment basés sur les logiciels d'aide à la prescription ;
- les bases de données à partir d'autres acteurs de la production de soins en ville (pharmacies d'officine, laboratoires de biologie et centres d'imagerie) ;
- les bases de données produites par le biais de dispositifs médicaux connectés d'automesures (rythme cardiaque, glycémie...) ou par la saisie directe par le patient d'informations dans des applications ;
- les bases de données issues de registres ou de cohortes académiques.

Le *Health Data Hub* et la base du système national des données de santé (SNDS)

Le *Health Data Hub* est la plateforme française des données de santé (<https://www.health-data-hub.fr/>) créée par la loi du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé. Ces données sont notamment celles de la base principale du système national des données de santé (SNDS créé par la loi de modernisation du système de santé français du 26 janvier 2016), mais également des établissements de santé, de Santé publique France, etc.^{6,7}.

Le SNDS rassemble et met à disposition des informations de santé collectées par des organismes publics. Cette fusion de plusieurs bases de données concerne à ce jour trois bases déjà existantes : le système national d'information inter-régimes de l'Assurance maladie (Sniiram) ; les données des hôpitaux et autres établissements de santé (programme de médicalisation des systèmes d'information – PMSI) ; les données statistiques relatives aux causes de décès (BCMD). Puis, lorsqu'elles seront constituées, le SNDS intégrera deux bases supplémentaires : les données « médico-sociales » des maisons départementales des personnes handicapées ; un échantillon représentatif des données de remboursement par bénéficiaires transmis par les mutuelles.⁸

⁶ <https://www.snds.gouv.fr/SNDS/Accueil>

⁷ Le texte fondateur du Sniiram est l'article L. 161-28-1, inscrit dans le Code de la sécurité sociale par la loi du 23 décembre 1998. L'entrepôt de données a été constitué en 2003, l'échantillon généraliste de bénéficiaires en 2005.

⁸ <https://www.snds.gouv.fr/SNDS/Composantes-du-SNDS>

Le SNDS contribue à la recherche, aux études, à l'évaluation et à l'innovation dans les domaines de la santé et de la prise en charge médico-sociale^{9,10} (165).

Avantages

Le recours aux sources de données secondaires peut en effet accélérer la disponibilité des résultats, limiter la duplication des collectes de données et donc *in fine* améliorer la collecte. Cela peut également favoriser les partenariats publics/privés.

Les données médico-administratives sont des données populationnelles et non un échantillonnage. Elles offrent des possibilités de suivi sur de longues périodes à l'échelle de larges populations, sans perdus de vue en cours de suivi, et une bonne homogénéité du codage. Le taux de données manquantes est faible.

Inconvénients

Le facteur limitant la mise en place d'études sur base de données est l'existence des données recherchées dans les bases existantes. Les items existants sont prédéfinis et donc peu flexibles. De nombreuses informations cliniques ne sont pas mesurées. Il n'y a pas de données concernant les facteurs de risque, la qualité de vie...

Par ailleurs, la mise à jour actuelle des données disponibles dans le SNDS est annuelle, ce qui peut retarder l'accès aux données de santé à des fins d'évaluation dans des temps compatibles avec la prise de décision.

Contraintes

La qualité des informations saisies dans les bases de données est essentielle : des indicateurs de qualité permettant de documenter et d'évaluer ces données sont nécessaires (166). L'optimisation de l'usage des bases disponibles est importante (amélioration des collectes d'information, registre déclaratif des bases de données de santé, outil de requêtage permettant aux acteurs de s'orienter entre les bases...). La transparence des méthodes utilisées pour analyser les données de santé est indispensable (167). Chaque fois que cela est possible, une étude multibase est à privilégier (par rapport à l'utilisation d'une seule base impliquant un risque de découverte fortuite) car elle permet de tester la cohérence entre les analyses et augmente la reproductibilité des résultats de l'étude.

Champs d'application

En France, la mise en place de la plateforme nationale des données de santé (*GIP Health Data Hub*) va développer l'utilisation des données en vie réelle **dans le respect du règlement général de la protection des données (RGPD)**. L'étape suivante sera le développement des données produites à l'occasion des soins fournis aux patients (168).

Dans les cas où le DM est connecté à des bases de données, celui-ci devient un outil de collecte d'information qui peut être utilisé à des fins d'études, dès lors que la réglementation en matière de

⁹ <https://documentation-snds.health-data-hub.fr/introduction/01-snds.html>

¹⁰ <https://www.cnil.fr/fr/snds-systeme-national-des-donnees-de-sante>

l'utilisation des données de santé est respectée¹¹. Le recueil de données directement auprès des patients est en pleine émergence, soit par le biais d'objets connectés d'automesure, soit par la saisie directe par le patient d'informations dans des applications ou plateformes dédiées.

Des initiatives pour faciliter l'identification des bases de données existantes sont en cours comme le programme de l'Alliance pour la recherche et l'innovation des industries de santé (ARIIS) et du Comité stratégique de filière des industries et technologies de santé¹² (118).

6.5. Comparaisons indirectes à partir d'études mono-bras [Nouveauté]

Contexte

L'essai de comparaison externe doit être réservé aux situations où la randomisation est irréalisable. En l'absence de comparaison directe, une comparaison indirecte, réalisée sur des bases méthodologiques définies et validées, peut être prise en compte. L'absence de comparaison directe au comparateur cliniquement pertinent devra être justifiée par l'industriel et pourra être acceptée dans certaines situations (169-171).

Principes

L'exploitation des études mono-bras pour démontrer le bénéfice d'un nouveau dispositif médical nécessite une comparaison implicite qui fera obligatoirement appel à une référence de comparaison externe (comparaison historique par exemple). La comparaison externe doit être formalisée dans le protocole de l'étude avec un choix *a priori* de la référence de comparaison ainsi que de la méthode de comparaison formelle : test par rapport à une norme, ajustement avec des données individuelles, groupe contrôle synthétique ou *Matching-Adjusted Indirect Comparisons* (MAIC) (172, 173). Les comparateurs externes possibles peuvent être une valeur de référence, fixée par une « exigence réglementaire » ou issue d'une revue systématique de la littérature à partir de données publiées ou de données individuelles d'une cohorte ou d'un essai randomisé.

Avantages

Les études mono-bras avec comparaisons externes peuvent être utilisées lorsque la population cible est restreinte et les inclusions difficiles.

Inconvénients

Pour être acceptables, ces études doivent pouvoir donner la garantie d'absence de facteurs de confusion résiduels. Cette comparaison est limitée par 1) le choix *post hoc* de la référence de la comparaison, 2) par le biais de confusion qui nécessite impérativement une approche d'ajustement, et 3) par les autres biais : de sélection, de mesure et liés aux données manquantes entre autres.

¹¹ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et hébergement des données de santé au sens de l'article L. 1111-8 du Code de la santé publique.

¹² <https://gt2.ariis.fr/sorienter-vers-les-donnees-pertinentes/la-forme-que-pourrait-prendre-un-tel-outil/>

Concernant le bras contrôle pour lequel une collecte de données rétrospective est mise en œuvre, les limites suivantes sont à prendre en compte :

- modification dans le temps de la stratégie thérapeutique modifiant l'histoire de la maladie ;
- augmentation de l'expérience de l'opérateur ainsi que de celle du patient dans le temps ;
- caractéristiques des patients différentes à l'inclusion entre un groupe contrôle historique et un groupe contemporain (151, 161).

L'expérience montre que trouver un comparateur adapté et prendre en compte tous les biais peut s'avérer bien plus compliqué et consommateur de temps que de faire un essai randomisé.

Contraintes

Afin de limiter les biais, il est nécessaire de définir *a priori* le contrôle externe et le justifier, c'est-à-dire garantir un choix non orienté de la valeur de référence du contrôle externe (sur la base d'une revue systématique des études de référence, en tenant compte de l'incertitude statistique de ces valeurs de référence en retenant la valeur la plus péjorative), d'identifier tous les facteurs de confusion potentiels (facteurs pronostiques des critères de jugement et modificateurs de l'effet du traitement) par une revue systématique, de prendre en compte ces facteurs de confusion et les biais potentiels, et de procéder à des analyses de sensibilité.

Champs d'application

Les données issues des études en vie réelle peuvent être également utilisées, sous conditions, comme bras contrôle dans une étude de comparaison indirecte, en particulier lorsqu'il existe un registre ou une collecte systématique des données. Cette utilisation est toutefois difficile lorsque les patients visés pour un DM nouveau diffèrent substantiellement de ceux des stratégies thérapeutiques de référence. Dans le domaine du DM, les études mono-bras peuvent permettre par exemple l'évaluation des évolutions d'un dispositif dont l'intérêt a initialement été démontré par un essai clinique randomisé, ou lorsque le nombre de patients disponibles est trop faible pour effectuer des essais cliniques randomisés (174).

6.6. Types d'analyse

6.6.1. Méthodes d'ajustement et score de propension [Nouveauté]

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 dans deux groupes, l'un recevant le 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 réduire les différences de pronostic entre les groupes comparés associées à la non-randomisation dans les études observationnelles (175).

Principe

Le score de propension peut être défini comme la probabilité pour un patient de recevoir le traitement A plutôt que le traitement B, en fonction de ses caractéristiques initiales observables (propension à recevoir un traitement en fonction de ses caractéristiques) (176).

L'objectif est d'équilibrer la distribution des covariables liées au pronostic du critère de jugement (par exemple âge, sexe, comorbidités, gravité de la maladie) entre les groupes, afin de neutraliser les facteurs de confusion (176, 177). Idéalement, l'ajustement devrait être réalisé sur tous les facteurs de confusion affectant le jeu de données et l'évaluation du critère principal.

La validité de cette méthode repose sur deux hypothèses : (1) l'hypothèse d'indépendance conditionnelle à des caractéristiques observables et (2) l'hypothèse de la condition de support commun. La première hypothèse signifie que l'appartenance au groupe de traitement ne doit pas dépendre du critère de jugement conditionnellement aux caractéristiques considérées dans le modèle. La seconde hypothèse suppose que les individus de chaque groupe comparé se ressemblent suffisamment pour que la comparaison ait un sens. Pour cela, il faut que la distribution du score de propension de chacun des deux groupes se recoupe suffisamment. L'absence de support commun interdirait toute conclusion quant à l'effet causal d'un traitement.

La méthode du score de propension comporte plusieurs étapes :

- (1) estimer le score de propension ;
- (2) vérifier l'existence d'un support commun. Il s'agit de vérifier l'existence d'un support commun (recouvrement ou *overlap*) afin de s'assurer qu'il est possible, pour chaque individu du groupe avec le traitement A, de trouver au moins un participant de l'autre groupe B ayant des caractéristiques voisines ;
- (3) vérifier la propriété équilibrante. La qualité de l'équilibre entre les groupes sur les variables associées au critère de jugement doit être vérifiée. Le but n'est pas de comparer les variables liées au critère de jugement entre les deux groupes par une analyse de comparaison, mais de vérifier qu'il n'y a pas de déséquilibre trop important sur ces variables. La vérification de la propriété équilibrante part du principe que le pourcentage de déséquilibre entre les groupes sur les variables associées au critère de jugement doit être suffisamment réduit. Le seuil de 10 % est utilisé pour juger d'un équilibre acceptable entre les deux groupes.

Il existe différentes méthodes pour prendre en compte le score de propension dans l'analyse (142, 178-182) :

- (1) modèle de régression avec ajustement par le score de propension ;
- (2) stratification sur le score de propension ;
- (3) appariement des sujets entre les groupes (*matching*) : les sujets sous traitement A doivent être appariés avec les sujets sous traitement B qui leur ressemblent. Cela consiste à former des paires de sujets qui ont les scores de propension les plus proches possibles. L'inconvénient de cette approche est qu'elle réduit la taille de l'échantillon au nombre de sujets appariés, entraînant une baisse de puissance, et que l'ensemble des sujets n'est pas inclus dans l'analyse. ;
- (4) pondération par l'inverse de la probabilité d'être traité (redressement ou IPTW)¹³ qui a pour but de créer une « pseudo-population » en attribuant un poids sur chaque sujet appartenant au groupe du

¹³ Il existe également une méthode de pondération par le ratio standardisé de mortalité (SMR).

traitement A et au groupe du traitement B. L'avantage de cette approche est qu'elle conserve l'effet sur l'ensemble de la population. C'est actuellement la méthode la plus utilisée.

La méthode par score de propension fonctionne mieux sous trois conditions (177) :

- lorsque l'évènement 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 prendre en compte les facteurs de confusion connus sur les variables mesurées en rendant les groupes comparables pour un score donné.

Inconvénients

L'utilisation du score de propension ne permet pas de garantir qu'il ne restera pas de la confusion, à la différence d'un essai randomisé bien conduit. On ne peut équilibrer les groupes que sur les facteurs de confusion connus et les variables recueillies et prises en compte dans l'estimation du score de propension (142, 177). Lorsqu'un grand nombre de données sont manquantes pour l'une des covariables, la validité de l'analyse du score de propension peut être difficilement interprétable (151). Il peut également exister un surajustement sur des facteurs qui ne sont pas des facteurs de confusion (*colliders*). Le risque de surajustement n'est pas propre au score de propension, c'est un problème avec tous les ajustements. Cependant, la réalisation d'analyses de sensibilité et le diagnostic d'un biais de confusion résiduel suivi d'un recalibrage peuvent permettre d'évaluer la robustesse des résultats (175, 183).

Contraintes

L'utilisation du score de propension repose sur trois hypothèses fortes qu'il faudra vérifier :

- 1) conditionnellement aux covariables, chaque sujet a la même probabilité d'être traité ;
- 2) toutes les variables de confusion sont identifiées, observées et mesurées ;
- 3) l'indépendance conditionnelle de l'attribution du traitement et des critères de jugement est vérifiée.

L'analyse par score de propension doit être planifiée. Toutes les variables pertinentes à collecter doivent être spécifiées dans le protocole. 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 (177).

Champs 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 ou de l'organisation des soins (154, 184, 185). Cette stratégie d'analyse pourrait être utilisée pour les dispositifs médicaux, en chirurgie et pour les techniques interventionnelles.

6.6.2. Variables instrumentales

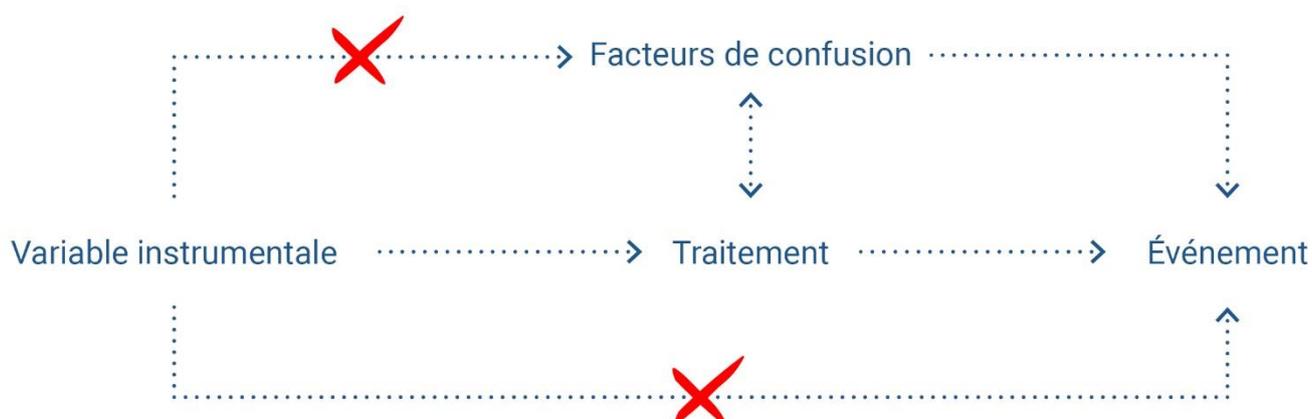
Contexte

Technique complexe à mettre en œuvre, les variables instrumentales sont utilisées en économétrie depuis des années, mais **leur utilisation dans le domaine de la santé est difficile** (186). 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 (187).

Principe

Une variable instrumentale est une variable fortement associée à l'indication du traitement (comme la préférence du prescripteur), mais non liée aux facteurs de risque de la survenue de l'évènement et n'affectant pas directement le critère de jugement (187) (figure 15).

Figure 15. Variable instrumentale, d'après (187)



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 à prendre en compte les facteurs de confusion inconnus (188).

L'approche par variable instrumentale commence par l'identification de la variable instrumentale (189).

Avantages

L'approche par variable instrumentale permet de réduire l'impact des facteurs de confusion dans les études observationnelles (189) et de prendre en compte les variables non mesurées (187, 188).

Inconvénients

L'estimation de l'effet du traitement est fondée sur des hypothèses fortes qui limitent l'utilisation de cette approche en pratique et qui sont également 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 du traitement est estimé pour les patients « marginaux », ceux qui sont observants (190). 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 la survenue de l'évènement dues à l'utilisation différentielle de ce traitement pourraient être très petites et, de ce fait, difficiles à évaluer (142).

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'évènement (191). D'autre part, un bon estimateur de variable instrumentale repose sur des effectifs larges (191). En cas de petits effectifs, les estimations sont incorrectes. Enfin, l'interprétation est difficile, en particulier quand l'effet traitement est hétérogène. Ces inconvénients impliquent que cette technique est difficile à mettre en œuvre.

Contraintes

Une bonne variable instrumentale doit répondre à trois critères (187) :

- avoir une relation forte avec l'indication du dispositif médical ; 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'évènement, mais uniquement à l'effet du dispositif médical.

Champs 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 des dispositifs médicaux et donc extrêmement **peu utilisées**.

7. En pratique

Le choix du plan d'évaluation clinique dépend totalement du contexte : l'action recherchée par l'utilisation du dispositif médical, l'existence d'alternatives, la taille de la population visée, etc. Ces différents paramètres doivent être pris en compte dans le choix de l'étude à mettre en place. À partir de ces éléments, le rationnel de décision devra être justifié par une argumentation scientifique explicite. C'est le point déterminant d'une stratégie de développement optimisée.

Pour toute étude, la qualité des données recueillies ou utilisées est un des prérequis indispensables pour que ses résultats soient pris en compte par la CNEDiMTS dans les évaluations des dispositifs médicaux et des produits de santé qu'elle examine.

Quel que soit le type d'étude choisi, le principe de transparence, qui est de rendre systématiquement publiques toutes les informations sur la recherche clinique tout en respectant la protection des données personnelles, doit être la règle. Il permet de renforcer la confiance des acteurs dans ces données, en mettant à disposition des informations sur le protocole et les éventuelles modifications qui y sont apportées, sur les sources de données et méthodes d'analyse (décrites suffisamment précisément pour permettre leur réplication), ainsi que sur les résultats de l'étude.

Ceci est d'autant plus important pour l'évaluation des innovations susceptibles d'apporter un réel bénéfice aux patients car **un continuum dans leur évaluation est souvent nécessaire entre les données expérimentales et les données en vie réelle.**

Les études contrôlées randomisées en double aveugle restent la référence de l'évaluation de tout produit de santé.

Si l'essai contrôlé randomisé ne peut être réalisé, alors le choix du type d'étude peut être guidé par différents éléments.

- Les caractéristiques du dispositif médical
 - Une évolution technologique importante et rapide
 - Essai avec randomisation adaptative
 - Méthodes bayésiennes
 - Éventuellement essais suiveurs ou *tracker trial design*
 - Une population cible faible
 - Méthodes bayésiennes
 - Essais en *cross-over* à la condition que la pathologie étudiée soit stable et que le critère de jugement puisse se répéter
 - Essais de type *Single Case Experimental Design* (SCED)
 - Essais séquentiels
 - Des événements indésirables potentiellement graves
 - Essais séquentiels

- L'acceptabilité médicale
 - La maîtrise de la technique influence le résultat là où les techniques sont très différentes
 - Essai reposant sur « l'expertise » ou *expertise-based randomised controlled trials*
 - Dans certaines circonstances, essai en *cluster*

→ L'acceptabilité par le patient

- Le comparateur est une technique invasive ou la technique est largement diffusée
- Plan expérimental de type Zelen ou *randomised consent trial*
- Essais fondés sur les préférences des patients ou étude observationnelle comparative si la randomisation est impossible

Table des annexes

Annexe 1.	Méthode de travail	63
Annexe 2.	Recherche documentaire	64
Annexe 3.	Liste des figures et des tableaux	71

Annexe 1. Méthode de travail

Les principales étapes sont décrites ci-dessous.

Phase de cadrage

Une feuille de route a été élaborée et a été soumise à la CNEDiMTS pour avis avant publication sur le site internet de la HAS.

Recueil et analyse des données

Une revue systématique de la littérature scientifique a été réalisée : la recherche documentaire mise en œuvre est détaillée en annexe 2.

Constitution du groupe de lecture et étape de relecture du guide

Un groupe de lecture, composé de méthodologistes, a été mis en place.

Le guide a été envoyé aux différents acteurs du groupe de lecture. À la suite de l'étape de relecture, les différents commentaires ont été analysés par le comité de pilotage pour modification éventuelle et validation du guide dans sa version définitive.

Examen par la CNEDiMTS

Les documents finalisés (guide et son rapport d'élaboration) ont été analysés et validés par la CNEDiMTS.

Diffusion des documents

Les documents finalisés (guide et son rapport d'élaboration) sont mis en ligne sur le site internet de la HAS.

Annexe 2. Recherche documentaire

Méthode

La recherche a porté sur les sujets et les types d'études définis en accord avec le chef de projet et a été limitée aux publications en langues anglaise et française.

La recherche initiale a porté sur la période de janvier 1999 ou janvier 2013 à octobre 2019. Une veille a ensuite été réalisée jusqu'à fin février 2021.

Les sources suivantes ont été interrogées :

- pour la littérature internationale : la base de données Medline ;
- la Cochrane Library ;
- les sites internet publiant des recommandations, des rapports d'évaluation technologique ou économique ;
- les sites internet des sociétés savantes compétentes dans le domaine étudié.

Cette recherche a été complétée par la bibliographie des experts et les références citées dans les documents analysés.

Bases de données bibliographiques

La stratégie de recherche dans les bases de données bibliographiques est construite en utilisant, pour chaque sujet, soit des termes issus de thésaurus (descripteurs), 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.

Tableau 1. Stratégie de recherche dans la base de données Medline (interface PubMed)

Type d'étude/sujet	Période	Nombre de références
Termes utilisés		
Choix méthodologiques pour l'évaluation des dispositifs médicaux		
Recommandations	01/2013 10/2019	-6
Étape 1	((Bayes Theorem OR Research Design ! OR Models, Statistical !)/de maj OR (bayesian OR Bayes OR study design OR methodological OR statistic* OR strateg*)/ti OR pre-market/ti,ab) AND ((Clinical Trials as Topic ! OR Comparative Effectiveness Research OR Meta-Analysis as Topic ! OR	

Practice Guidelines as Topic OR Device Approval ! OR Product Surveillance, Postmarketing ! OR Reproducibility of Results ! OR Sample Size)/de OR premarket/ti,ab) AND ((Equipment Safety OR Equipment and Supplies ! OR Biomedical Technology !)/de OR device*/ti)

ET

Étape 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

Méta-analyses, revues systématiques	01/2013	-73
	10/2019	

Étape 1

ET

Étape 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 OR pooled analysis)/ti OR (Meta-Analysis OR Systematic Review)/pt OR Cochrane Database Syst Rev/so

Essais contrôlés randomisés	01/2013	-46
	10/2019	

Étape 1

ET

Étape 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	01/2013	-67
	10/2019	

Étape 1

ET

Étape 5 random*/tia,b 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 types d'études	01/2013	-116
	10/2019	

Étape 1

SAUF Étape 2 OU Étape 3 OR Étape 4 OU Étape 5

Dispositifs médicaux et évaluation des technologies biomédicales

Tout type d'étude 01/2013 -45
10/2019

Étape 6 (Technology Assessment, Biomedical/methods/de OR technology assessment/ti) AND (Equipment and Supplies !/de maj OR medical device*/ti)

Évaluation des dispositifs et bases de données médico-administratives

Recommandations 01/2013 -2
10/2019

Étape 7 (Equipment Safety OR Equipment and Supplies ! OR Biomedical Technology !)/de maj OR device*/ti

AND

(Clinical Trials as Topic! OR Meta-Analysis as Topic ! OR Guidelines as Topic ! OR Health Services Research ! OR Research Design ! OR Process Assessment, Health Care OR Device Approval ! OR Product Surveillance, Postmarketing ! OR Probability !)/de maj OR (premarket OR postmarket OR design OR designs)/ti OR propensity score/ti,ab)

AND

((Registries ! OR Records ! OR Databases as Topic !)/de maj OR (database* OR register OR registry OR registries OR health record* OR medical record*)/ti OR (national database* OR administrative database* OR public regulatory database* OR public database*)/ti,ab

ET Étape 2

Méta-analyses, revues systématiques 01/2013 -9
10/2019

Étape 7 ET Étape 3

Essais contrôlés randomisés 01/2013 -16
10/2019

Étape 7 ET Étape 4

Essais contrôlés 01/2013 -90
10/2019

Étape 7 ET Étape 5

Études observationnelles 01/2013 -46
10/2019

Étape 7

ET

Étape 8 (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 OR Observational Study/pt

Autres types d'études	01/2013 10/2019	-74
-----------------------	--------------------	-----

Étape 7

SAUF Étape 2 OU Étape 3 OU Étape 4 OU Étape 5 OU Étape 8

Dispositifs médicaux et méthodes *in silico*

Tout type d'étude	01/2013 10/2019	-66
-------------------	--------------------	-----

Étape 9 (Computer Simulation !/de maj OR (computer simulation* OR computerized model* OR computer model* OR in silico*/ti,ab OR in silico*/ot)

AND

((Equipment Safety OR Equipment and Supplies ! OR Bio-medical Technology!)/de OR device*/ti,ab)

AND

(Clinical Trials as Topic !/de OR Clinical Trial !/pt OR (clinical trial* OR clinical stud*/ti)

Single-case experimental design : méthodologie

Tout type d'étude	01/2009 06/2019	-104
-------------------	--------------------	------

Étape 10 (single-case experimental design* OR single-case design* OR SCED OR N-of-1* OR small population group trial* OR single subject experimental design OR single subject research design* OR small N design* OR multiple-case design* OR single-case design* OR single-systems design*/ti,ab OR (single-case experimental design* OR single-case design* OR SCED OR N-of-1* OR small population group trial* OR single subject experimental design OR single subject research design* OR small N design* OR multiple-case design* OR single-case design* OR single-systems design*/ot OR (small population*/ti AND clinical trial*/ti)

ET

Étape 11 (assess* OR analysis* OR effectiveness OR methods OR implementation OR guide OR bias OR evidence-based)/ti OR

method/ti,ab OR (Research Design ! OR Data Interpretation, Statistical)/de maj OR (Bias ! OR Evidence-Based Practice !/methods)/de

Single-case experimental design et dispositifs médicaux

Tout type d'étude	01/1999	-53
	06/2019	

Étape 12 Equipment and Supplies !/de OR medical device*/ti,ab

ET Étape 10

Single-case experimental design et médecine physique et de réadaptation

Tout type d'étude	01/1999	-48
	06/2019	

Étape 13 Physical and Rehabilitation Medicine !/de OR (neurorehabilitation OR rehabilitation)/ti

ET Étape 10

de : descripteur ; de maj : descripteur majoré ; * : troncature ; ti : titre ; ab : résumé ; pt : type de publication ; ! : explosion du terme générique ; so : titre du journal ; ot : mots clés auteurs

Sites consultés

Dernière consultation : octobre 2019

Bibliothèque médicale Lemanissier

Catalogue et index des sites médicaux francophones – CISMef

Comité d'évaluation et de diffusion des innovations technologiques – CEDIT

Adelaide Health Technology Assessment – AHTA

Agency for Care Effectiveness

Agency for Healthcare Research and Quality – AHRQ

Alberta Health – HTA provincial reviews

Alberta Medical Association

Allied Health Evidence

American College of Physicians – ACP

Australian Clinical Practice Guideline

Australia and New Zealand Horizon Scanning Network

BMJ Best Practice

British Columbia guidelines

California Technology Assessment Forum
Campbell Collaboration
Canadian Agency for Drugs and Technologies in Health – CADTH
Canadian Task Force on Preventive Health Care
Centers for Disease Control and Prevention – CDC
Centre fédéral d’expertise des soins de santé – KCE
Centre for Clinical Effectiveness – CCE
Centre for Effective Practice
Centre for Reviews and Dissemination databases
CMA Infobase
Cochrane Library
Guidelines International Network – GIN
Health Services Technology Assessment Text – HSTAT
Institute for Clinical and Economic Review – ICER
Institute for Clinical Evaluative Sciences – ICES
Institute for Clinical Systems Improvement – ICSI
Institute for Health Economics Alberta – IHE
Institut national d’excellence en santé et en services sociaux – INESSS
International Network of Agencies for Health Technology Assessment – INAHTA
McGill University Health Centre
Malaysian Health Technology Assessment Section
Medical Services Advisory Committee – MSAC
National Coordinating Centre for Health Technology Assessment – NCCHTA
National Health and Medical Research Council – NHMRC
National Health Services Evidence
National Health Services Innovation Observatory
National Institute for Health and Clinical Excellence – NICE
New South Wales Agency for Clinical Innovation
New Zealand Guidelines Group – NZGG
Ontario Health Technology Advisory Committee – OHTAC
Public Health Agency of Canada
Scottish Health Technologies Group
Scottish Intercollegiate Guidelines Network – SIGN
Singapore Ministry of Health
Tripdatabase
U.S. Food and Drug Administration

U.S. Preventive Services Task Force

Veterans affairs, Dep. Of Defense Clinical practice guidelines

Veterans Affairs Evidence-based Synthesis Program

Veille

En complément, une veille a été réalisée jusqu'à fin février 2021 dans Medline, sur la base des équations du tableau 1 et sur les sites indiqués ci-dessus.

Annexe 3. Liste des figures et des tableaux

Table des figures

Figure 1. Arbre décisionnel pour limiter les biais de mesure dans les essais randomisés	18
Figure 2. Prérandomisation d'après Zelen, cas 1 (45)	19
Figure 3. Prérandomisation d'après Zelen, cas 2 (45)	20
Figure 4. Double consentement d'après Zelen, cas 3 (47)	20
Figure 5. Schéma représentant la randomisation dans les essais contrôlés randomisés basés sur l'expertise, d'après (39)	23
Figure 6. Schéma représentant les essais en <i>cluster</i> , d'après (61)	27
Figure 7. Principe du <i>cross-over</i>	30
Figure 8. Cas particulier du <i>cross-over</i> lorsque deux traitements sont appliqués simultanément	31
Figure 9. Choix du type de SCED, d'après (72)	33
Figure 10. Règles de décision graphique pour le test triangulaire (plan fermé), d'après (107, 109)	39
Figure 11. Règle de décision graphique pour le test séquentiel de rapport des probabilités (plan ouvert)	39
Figure 12. Pourquoi mettre en place une étude en vie réelle, d'après le guide de la HAS « Études en vie réelle pour l'évaluation des médicaments et des dispositifs médicaux »	47
Figure 13. Exemple d'étude observationnelle prospective comparant la survenue d'infections postopératoires selon le type de suture utilisé	48
Figure 14. Principe des études basées sur les préférences ou des <i>comprehensive cohort studies</i> , d'après (163)	51
Figure 15. Variable instrumentale, d'après (187)	58

Table des tableaux

Tableau 1. Exemples d'impossibilité de réalisation de l'étude en aveugle, d'après (30)	13
Tableau 2. Exemples de placebos utilisés en chirurgie et pour les dispositifs médicaux, d'après (24) et (30)	14

Références bibliographiques

1. Haute Autorité de Santé. Parcours du dispositif médical en France. Guide mis à jour en octobre 2020. Saint-Denis La Plaine: HAS; 2017.
https://www.has-sante.fr/upload/docs/application/pdf/2009-12/guide_pratique_dm.pdf
2. Haute Autorité de Santé. Principes d'évaluation de la CNEDiMTS relatifs aux dispositifs médicaux à usage individuel en vue de leur accès au remboursement. Évaluation des dispositifs médicaux. HAS: Saint-Denis La Plaine; 2019.
https://www.has-sante.fr/upload/docs/application/pdf/2017-11/principes_devaluation_de_la_cnedimts-v4-161117.pdf
3. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898.
<http://dx.doi.org/10.1136/bmj.l4898>
4. Centre fédéral 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: KCE; 2006.
https://kce.fgov.be/sites/default/files/atoms/files/d20061027351_0.pdf
5. Simpson S, Hiller J, Gutierrez-Ibarluzea I, Kearney B, Norderhaug I, Fay A, *et al.* A toolkit for the identification and assessment of new and emerging health technologies. Birmingham: EuroScan; 2009.
https://www.researchgate.net/profile/Inaki_Gutierrez-Ibarluzea/publication/280877428_A_toolkit_for_the_identification_and_assessment_of_new_and_emerging_health_technologies/links/55c9c3c608aebc967df93863/A-toolkit-for-the-identification-and-assessment-of-new-and-emerging-health-technologies.pdf?origin=publication_detail
6. Bernard A. Le développement clinique des nouveaux dispositifs médicaux. Paris: Académie Nationale de Chirurgie; 2012.
7. National Institute for Health and Clinical Excellence. Medical technologies evaluation programme methods guide . Process and methods London: NICE; 2017.
<https://www.nice.org.uk/process/pmg33/resources/medical-technologies-evaluation-programme-methods-guide-pdf-72286774205893>
8. Centre fédéral d'expertise des soins en santé. Towards a guided and phased introduction of high-risk medical devices in Belgium. Brussels: KCE; 2015.
https://kce.fgov.be/sites/default/files/atoms/files/KCE249_High-risk%20medical%20devices_Report_0.pdf
9. Schnell-Inderst P, Hunger T, Conrads-Frank A, Arvandi M, Siebert U. Recommendations for primary studies evaluating therapeutic medical devices were identified and systematically reported through reviewing existing guidance. *J Clin Epidemiol* 2018;94:46-58.
<http://dx.doi.org/10.1016/j.jclinepi.2017.10.007>
10. Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA* 2002;288(3):358-62.
<http://dx.doi.org/10.1001/jama.288.3.358>
11. Haute Autorité de Santé. Spécificités méthodologiques d'évaluation clinique d'un dispositif médical connecté (DMC). Saint-Denis La Plaine: HAS; 2019.
https://www.has-sante.fr/upload/docs/application/pdf/2019-02/rapport_methodologiques_devaluation_clinique_dun_dmc.pdf
12. Bernard A, Vicaut E. Quelles études pour évaluer des nouveaux dispositifs médicaux ? *J Chir* 2009;146(2):129-35.
13. Bernard A, Vicaut E. Dispositifs médicaux. Première partie : de quels types d'études avons-nous besoin ? *Médecine* 2008;4(10):464-7.
14. 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.
15. 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).
16. Sauerland S, Fujita-Rohwerder N, Zens Y, Molnar S. Premarket evaluation of medical devices: a cross-sectional analysis of clinical studies submitted to a German ethics committee. *BMJ open* 2019;9(2):e027041.
<http://dx.doi.org/10.1136/bmjopen-2018-027041>
17. 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.

18. Weil RJ. The future of surgical research. *PLoS Med* 2004;1(1):e13.
19. Li H, Yue LQ. Statistical and regulatory issues in nonrandomized medical device clinical studies. *J Biopharm Stat* 2008;18(1):20-30.
20. 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.
21. Zarin DA, Goodman SN, Kimmelman J. Harms From Uninformative Clinical Trials. *JAMA* 2019;322(9):813-4.
<http://dx.doi.org/10.1001/jama.2019.9892>
22. Boutron I, Guttet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 2007;4(2).
23. 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.
24. 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.
25. Heckerling PS. The ethics of single blind trials. *IRB* 2005;27(4):12-6.
26. Barker JP, Simon SD, Dubin J. The methodology of clinical studies used by the FDA for approval of high-risk orthopaedic devices. *J Bone Joint Surg Am* 2017;99(9):711-9.
<http://dx.doi.org/10.2106/jbjs.16.00403>
27. Rathi VK, Wang B, Ross JS, Downing NS, Kesselheim AS, Gray ST. Clinical evidence supporting US food and drug administration premarket approval of high-risk otolaryngologic devices, 2000-2014. *Otolaryngol Head Neck Surg* 2017;156(2):285-8.
<http://dx.doi.org/10.1177/0194599816684094>
28. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. Systematic reviews in health care. *BMJ* 2001;323(7303):42-6.
29. Kinnaert P. La chirurgie factuelle : aspects éthiques et méthodologiques. *Rev Med Brux* 2006;27:451-8.
30. Kinnaert P. Placebo et effet placebo (deuxième partie) : aspects éthiques. *Rev Med Brux* 2007;28:39-44.
31. Byrne RA, Capodanno D, Mahfoud F, Fajadet J, Windecker S, Jüni P, *et al.* Evaluating the importance of sham-controlled trials in the investigation of medical devices in interventional cardiology. *EuroIntervention* 2018;14(6):708-15.
<http://dx.doi.org/10.4244/eij-d-18-00481>
32. Lauder L, da Costa BR, Ewen S, Scholz SS, Wijns W, Lüscher TF, *et al.* Randomized trials of invasive cardiovascular interventions that include a placebo control: a systematic review and meta-analysis. *Eur Heart J* 2020;41(27):2556-69.
<http://dx.doi.org/10.1093/eurheartj/ehaa495>
33. Rohrer R. Sham surgery. *Virtual Mentor* 14(3):227-31.
<http://dx.doi.org/10.1001/virtualmentor.2012.14.3.stas2-1203>
34. Cohen PJ. Failure to Conduct a Placebo-Controlled Trial May Be Unethical. *The American Journal of Bioethics* 2002;2(2).
35. DeMaria A. How important are placebo controls in clinical trials of interventional procedures? *Eur Heart J* 2020;41(27):2569-70.
<http://dx.doi.org/10.1093/eurheartj/ehaa558>
36. Macklin R. The ethical problems with sham surgery in clinical research. *N Engl J Med* 1999;341(13):992-6.
<http://dx.doi.org/10.1056/nejm199909233411312>
37. Niemansburg SL, van Delden JJ, Dhert WJ, Bredenoord AL. Reconsidering the ethics of sham interventions in an era of emerging technologies. *Surgery* 2015;157(4):801-10.
<http://dx.doi.org/10.1016/j.surg.2014.12.001>
38. Boutron I, Estellat C, Guttet 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.

39. Rudicel S, Esdaile J. The randomized clinical trial in orthopaedics: obligation or option? *J Bone Joint Surg Am* 1985;67(8):1284-93.
40. Campbell G. Statistics in the world of medical devices: the contrast with pharmaceuticals. *J Biopharm Stat* 2008;18(1):4-19.
41. 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.
42. Stone GW, Grines CL, O'Neill WW. Primary coronary angioplasty versus thrombolysis. *N Engl J Med* 1997;337(16):1168-9.
43. 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.
44. Ravaud P. Evaluation des traitements non médicamenteux. 4^{ème} conférence d'épidémiologie clinique. 26-28 mai 2010. Paris; 2011.
45. Zelen M. A new design for randomized clinical trials. *N Engl J Med* 1979;300(22):1242-5.
46. Blondel B, Goffinet F, Bréart. Evaluation des soins en obstétrique. Paris: Masson; 2001.
47. Zelen M. Innovations in the design of clinical trials in breast cancer. *Breast Cancer Res Treat* 1983;3(2):137-42.
48. Zelen M. Strategy and alternate randomized designs in cancer clinical trials. *Cancer Treat Rep* 1982;66(5):1095-100.
49. Zelen M. Randomized consent designs for clinical trials: an update. *Stat Med* 1990;9(6):645-56.
50. Horwitz RI, Feinstein AR. Advantages and drawbacks of the Zelen design for randomized clinical trials. *J Clin Pharmacol* 1980;20(7):425-7.
51. Flory JH, Mushlin AI, Goodman ZI. Proposals to conduct randomized controlled trials without informed consent: a narrative review. *J Gen Intern Med* 2016;31(12):1511-8.
<http://dx.doi.org/10.1007/s11606-016-3780-5>
52. 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.
53. 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.
54. DeTurrís SV, Cacchione RN, Mungara A, Pecoraro A, Ferzli GS. Laparoscopic herniorrhaphy: beyond the learning curve. *J Am Coll Surg* 2002;194(1):65-73.
55. Menon VS, Manson JMCK, Baxter JN. Laparoscopic fundoplication: learning curve and patient satisfaction. *Ann R Coll Surg Engl* 2003;85(1):10-3.
56. 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.
57. Cook JA, Elders A, Boachie C, Bassinga T, Fraser C, Altman DG, *et al.* A systematic review of the use of an expertise-based randomised controlled trial design. *Trials* 2015;16:241.
<http://dx.doi.org/10.1186/s13063-015-0739-5>
58. Cook JA, Campbell MK, Gillies K, Skea Z. Surgeons' and methodologists' perceptions of utilising an expertise-based randomised controlled trial design: a qualitative study. *Trials* 2018;19(1):478.
<http://dx.doi.org/10.1186/s13063-018-2832-z>
59. Lilford RJ, Braunholtz DA, Greenhalgh R, Edwards SJL. Trials and fast changing technologies: the case for tracker studies. *BMJ* 2000;320(7226):43-6.
60. Eldridge S, Campbell M, Campbell M, Drahota-Towns A, Giraudeau B, Reeves B, *et al.* Revised Cochrane risk of bias tool for randomized trials (RoB 2.0). Additional considerations for cluster-randomized trials (RoB 2 CRT) ; 2020.
https://drive.google.com/file/d/1J4okZ1zMWzF2LSe8Y_3XTYghAZfUset/view

61. Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomised trials. *BMC Med Res Methodol* 2005;5:10. [http://dx.doi.org/10.1016/s0140-6736\(09\)61459-8](http://dx.doi.org/10.1016/s0140-6736(09)61459-8)
62. Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. *PLoS Med* 2009;6(5). <http://dx.doi.org/10.1016/j.rehab.2017.12.002>
63. Ford I, Norrie J. Pragmatic trials. *N Engl J Med* 2016;375(5):454-63. <http://dx.doi.org/10.1056/NEJMra1510059>
64. Haute Autorité de Santé. Méthodes quantitatives pour évaluer les interventions visant à améliorer les pratiques. Guide méthodologique. Saint-Denis La Plaine: HAS; 2007. https://has-sante.fr/upload/docs/application/pdf/eval_interventions_ameliorer_pratiques_guide.pdf
65. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000;9(2):81-94.
66. 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. <http://dx.doi.org/10.1080/09602011.2013.824383>
67. Caille A, Leyrat C, Giraudeau B. L'essai randomisé en cluster. *Ann Dermatol Venereol* 2014;141(2):144-6. <http://dx.doi.org/10.1016/j.annder.2013.07.024>
68. Higgins J, Li T, Sterne J. Revised Cochrane risk of bias tool for randomized trials (RoB 2). Additional considerations for crossover trials. Preliminary tool version ; 2020. <https://drive.google.com/file/d/18Ek-uW8HYQsUja8Lakp1yOhoFk0EMfPO/view>
69. Elbourne DR, Altman DG, Higgins JP, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol* 2002;31(1):140-9.
70. Hills M, Armitage P. The two-period cross-over clinical trial. *Br J Clin Pharmacol* 1979;8(1):7-20.
71. Hejblum G, Chalumeau-Lemoine L, Ioos V, Boëlle PY, Salomon L, Simon T, *et al.* Comparison of routine and on-demand prescription of chest radiographs in mechanically ventilated adults: a multicentre, cluster-randomised, two-period crossover study. *Lancet* 2009;374(9702):1687-93. <http://dx.doi.org/10.1093/bib/bby043>
72. Krasny-Pacini A, Evans J. Single-case experimental designs to assess intervention effectiveness in rehabilitation: a practical guide. *Ann Phys Rehabil Med* 2018;61(3):164-79. <http://dx.doi.org/10.1016/j.rehab.2017.12.002>
73. Kratochwill TR, Hitchcock J, Horner RH, Levin JR, Odom SL, Rindskopf DM, *et al.* Single-case design technical documentation. Version 1.0 (Pilot) ; 2010. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_scd.pdf
74. Lobo MA, Moeyaert M, Baraldi Cunha A, Babik I. Single-case design, analysis, and quality assessment for intervention research. *JNPT* 2017;41(3):187-97. <http://dx.doi.org/10.1097/npt.000000000000187>
75. Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, *et al.* Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychol Rehabil* 2013;23(5):619-38. <http://dx.doi.org/10.1080/09602011.2013.824383>
76. Viceconti M, Cobelli C, Haddad T, Himes A, Kovatchev B, Palmer M. In silico assessment of biomedical products: The conundrum of rare but not so rare events in two case studies. *Proc Inst Mech Eng [H]* 2017;231(5):455-66. <http://dx.doi.org/10.1177/0954411917702931>
77. Ko H, Dufull S. Simulation for designing clinical trials : a pharmacokinetic-pharmacodynamic modeling perspective. New-York: Marcel Dekker; 2002.
78. Holford N, Hale M, Ko H, Steimer J-L, Sheiner L, Peck C. Simulation in drug development: good practices. Washington: CDDS; 1999. <http://holford.fmhs.auckland.ac.nz/docs/simulation-in-drug-development-good-practices.pdf>
79. Faris O, Shuren J. An FDA viewpoint on unique considerations for medical-device clinical trials. *N Engl J Med* 2017;376(14):1350-7. <http://dx.doi.org/10.1056/NEJMra1512592>
80. Girard P, Cucherat M, Guez D, les participants de la Table Ronde No 2 de Giens XIX. Simulation des essais cliniques dans le développement des médicaments. *Thérapie* 2004;59(3):287-95.
81. Pappalardo F, Russo G, Tshinanu FM, Viceconti M. In silico clinical trials: concepts and early adoptions. *Brief Bioinform* 2018. <http://dx.doi.org/10.1093/bib/bby043>

82. Ezzati A, Lipton RB. Machine learning predictive models can improve efficacy of clinical trials for alzheimer's disease. *J Alzheimers Dis* 2020;74(1):55-63.
<http://dx.doi.org/10.3233/jad-190822>
83. Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent progress. *Precision Oncology* 2020;4(1):19.
<http://dx.doi.org/10.1038/s41698-020-0122-1>
84. Spiegelhalter DJ, Abrams KR, Myles P. Bayesian approaches to clinical trials and health-care evaluation. London: John Wiley & Sons; 2004.
<https://onlinelibrary.wiley.com/doi/book/10.1002/0470092602>
85. Center for Devices and Radiological Health, Center for Biologics Evaluation and Research, Campbell G. Guidance for industry and FDA staff. Guidance for the use of bayesian statistics in medical device clinical trials. Rockville: FDA; 2010.
<https://www.fda.gov/media/71512/download>
86. Food and Drug Administration. Guidance for the use of bayesian statistics in medical device clinical trials. Guidance for industry and FDA staff. Silver Spring: FDA; 2010.
<https://www.fda.gov/media/71512/download>
87. 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.
88. Reardon MJ, Van Mieghem NM, Popma JJ, Kleiman NS, Søndergaard L, Mumtaz M, *et al.* Surgical or transcatheter aortic-valve replacement in intermediate-risk patients. *N Engl J Med* 2017;376(14):1321-31.
<http://dx.doi.org/10.1056/NEJMoa1700456>
89. Haute Autorité de Santé. Avis de la CNEDiMITS 10 juillet 2018. COREVALVE EVOLUT PRO avec cathéter de pose EnVeoR, bioprothèse valvulaire aortique implantée par voie artérielle transcutanée. Saint-Denis La Plaine: HAS; 2020.
[https://www.has-sante.fr/upload/docs/evamed/CNEDiMITS-6224_COREVALVE%20EVOLUT%20PRO_21_juillet_2020_\(6224\)_avis.pdf](https://www.has-sante.fr/upload/docs/evamed/CNEDiMITS-6224_COREVALVE%20EVOLUT%20PRO_21_juillet_2020_(6224)_avis.pdf)
90. Popma JJ, Deeb GM, Yakubov SJ, Mumtaz M, Gada H, O'Hair D, *et al.* Transcatheter aortic-valve replacement with a self-expanding valve in low-risk patients. *N Engl J Med* 2019;380(18):1706-15.
<http://dx.doi.org/10.1056/NEJMoa1816885>
91. Blanke P, Leipsic JA, Popma JJ, Yakubov SJ, Deeb GM, Gada H, *et al.* Bioprosthetic aortic valve leaflet thickening in the evolut low risk sub-study. *JACC* 2020;75(19):2430-42.
<http://dx.doi.org/https://doi.org/10.1016/j.jacc.2020.03.022>
92. Haute Autorité de Santé. Avis de la CNEDiMITS du 2 février 2021. COREVALVE EVOLUT R avec système de pose EnVeo PRO, bioprothèses valvulaires aortiques implantées par voie artérielle transcutanée. Saint-Denis La Plaine: HAS.
[https://has-sante.fr/upload/docs/evamed/CNEDiMITS-6421_COREVALVE_EVOLUT_R_02_f%C3%A9vrier_2021_\(6421\)_avis.pdf](https://has-sante.fr/upload/docs/evamed/CNEDiMITS-6421_COREVALVE_EVOLUT_R_02_f%C3%A9vrier_2021_(6421)_avis.pdf)
93. Psioda MA, Ibrahim JG. Bayesian clinical trial design using historical data that inform the treatment effect. *Biostatistics* 2019;20(3):400-15.
<http://dx.doi.org/10.1093/biostatistics/kxy009>
94. Psioda MA, Soukup M, Ibrahim JG. A practical Bayesian adaptive design incorporating data from historical controls. *Stat Med* 2018;37(27):4054-70.
<http://dx.doi.org/10.1002/sim.7897>
95. 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.
96. Wassmer G. Basic concepts of group sequential and adaptative group sequential test procedures. *Stat Pap* 2000;41:253-79.
97. Armitage P. Sequential medical trials. New York: Wiley; 1975.
98. 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.
99. Pocock S. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977;64(2):191-9.
100. Pocock SJ. Interim analyses for randomized clinical trials: the group sequential approach. *Biometrics* 1982;38(1):153-62.

101. Pocock SJ, Hughes M. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials* 1989;10(4 Suppl):209S-21S.
102. Lan GKK, Simon R, Halperin M. Stochastically curtailed tests in long-term clinical trials. *Sequential Analysis* 1982;1(3):207-19.
103. Falissard B, Lellouch J. Some extensions to a new approach for interim analysis in clinical trials. *Stat Med* 1991;10(6):949-55.
104. Falissard B, Lellouch J. A new procedure for group sequential analysis in clinical trials. *Biometrics* 1992;48(2):373-88.
105. Whitehead J, Stratton I. Group sequential clinical trials with triangular continuation regions. *Biometrics* 1983;39(1):227-36.
106. DeMets DL, Ware JH. Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika* 1980;67(3):651-60.
107. 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.
108. Clayton DG. Ethically optimised designs. *Br J Clin Pharmacol* 1982;13(4):469-80.
109. 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.
110. Sebillé V, Bellissant E. Comparison of four sequential methods allowing for early stopping of comparative clinical trials. *Clin Sci* 2000;98(5):569-78.
111. Coad DS, Ivanova A. The use of the triangular test with response-adaptive treatment allocation. *Stat Med* 2005;24(10):1483-93.
112. Bauer P, Brannath W. The advantages and disadvantages of adaptive designs for clinical trials. *Drug Discov Today* 2004;9(8):351-7.
113. Weinstein MC. Allocation of subjects in medical experiments. *N Engl J Med* 1974;291(24):1278-85.
114. Day N. A comparison of some sequential designs. *Biometrika* 1969;56:301-11.
115. Lellouch J, Schwartz D. L'essai thérapeutique : éthique individuelle ou éthique collective. *Rev Inst Int Stat* 1971;39:127-36.
116. Simon R. Adaptive treatment assignment methods and clinical trials. *Biometrics* 1977;33(4):743-9.
117. Hung HMJ. Considerations in adapting clinical trial design. *J Formos Med Assoc* 2008;107(12 Suppl):S14-S8.
118. Levy-Bachelot L, Mollo L, Josseran A, Bacry E, Combes S, Berille J, *et al.* Programme « Intelligence Artificielle et Santé » du comité stratégique de filière industries et technologies de santé. Recommandations du groupe de travail « usage des données en vie réelle dans la prise de décision ». Paris: Conseil national de l'industrie; ARIIS.
https://qt2.ariis.fr/wp-content/uploads/2021/03/Recommandations-du-GT-Usage-des-donnees-en-vie-reelle-dans-la-prise-de-decision_vf.pdf
119. Visvanathan K, Levit LA, Raghavan D, Hudis CA, Wong S, Dueck A, *et al.* Untapped potential of observational research to inform clinical decision making: American Society of Clinical Oncology research statement. *J Clin Oncol* 2017;35(16):1845-54.
<http://dx.doi.org/10.1200/jco.2017.72.6414>
120. Frieden TR. Evidence for health decision making—beyond randomized, controlled trials. *N Engl J Med* 2017;377(5):465-75.
<http://dx.doi.org/10.1056/NEJMra1614394>
121. Bégaud B, Polton D, von Lennep F. Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé. L'exemple du médicament. Paris: Ministère des solidarités et de la santé; 2017.
122. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, *et al.* Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on

- real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf* 2017;26(9):1033-9.
<http://dx.doi.org/10.1002/pds.4297>
123. Dahabreh IJ, Robins JM, Hernán MA. Benchmarking observational methods by comparing randomized trials and their emulations. *Epidemiology* 2020;31(5):614-9.
<http://dx.doi.org/10.1097/ede.0000000000001231>
124. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. *Nat Med* 2019;25(10):1601-6.
<http://dx.doi.org/10.1038/s41591-019-0597-x>
125. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol* 2016;79:70-5.
<http://dx.doi.org/10.1016/j.jclinepi.2016.04.014>
126. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183(8):758-64.
<http://dx.doi.org/10.1093/aje/kwv254>
127. Danaei G, Rodríguez LA, Cantero OF, Logan R, Hernán MA. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res* 2013;22(1):70-96.
<http://dx.doi.org/10.1177/0962280211403603>
128. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Systematic Reviews* 2014;Issue 4:MR000034.
<http://dx.doi.org/10.1002/14651858.MR000034.pub2>
129. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, *et al.* Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286(7):821-30.
<http://dx.doi.org/10.1001/jama.286.7.821>
130. Grodstein F, Clarkson TB, Manson JE. Understanding the divergent data on postmenopausal hormone therapy. *N Engl J Med* 2003;348(7):645-50.
<http://dx.doi.org/10.1056/NEJMs022365>
131. Pongiglione B, Torbica A, Blommestein H, de Groot S, Ciani O, Walker S, *et al.* Do existing real-world data sources generate suitable evidence for the HTA of medical devices in Europe? Mapping and critical appraisal. *Int J Technol Assess Health Care* 2021;37(1):e62.
<http://dx.doi.org/10.1017/s0266462321000301>
132. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandembroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg* 2014;12(12):1495-9.
<http://dx.doi.org/10.1016/j.ijsu.2014.07.013>
133. Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, *et al.* STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ* 2021;372:m4856.
<http://dx.doi.org/10.1136/bmj.m4856>
134. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
<http://dx.doi.org/10.1136/bmj.i4919>
135. Health Quality Ontario. Health technology assessments. Methods and process guide, version 2.0. Toronto: HQO; 2018.
<https://www.hqontario.ca/Portals/0/documents/evidence/reports/hta-methods-and-process-guide-en.pdf>
136. Food and Drug Administration. Use of real-world evidence to support regulatory decision-making for medical device. Guidance for industry and Food and Drug Administration staff. Silver Spring: FDA; 2017.
<https://www.fda.gov/files/medical%20devices/published/Use-of-Real-World-Evidence-to-Support-Regulatory-Decision-Making-for-Medical-Devices---Guidance-for-Industry-and-Food-and-Drug-Administration-Staff.pdf>
137. Salines G, de Launay C. Les cohortes. Intérêt, rôle et position de l'InVS. Saint-Maurice: InVS; 2010.
<https://www.santepubliquefrance.fr/docs/les-cohortes.-interet-role-et-position-de-l-invs>
138. Agency for Healthcare Research and Quality. Registries for evaluating patient outcomes: a user's guide. 4th ed. Rockville: AHRQ; 2020.
<https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/registries-evaluating-patient-outcomes-4th-edition.pdf>
139. Sackett DL. Why did the randomized clinical trial become the primary focus of my career? *Value Health* 2015;18(5):550-2.
<http://dx.doi.org/10.1016/j.jval.2015.04.001>

140. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, *et al.* Risks and benefits of estrogen plus progestin in healthy postmenopausal women. *JAMA* 2002;288(3):321-33.
<http://dx.doi.org/10.1001/jama.288.3.321>
141. Vandembroucke JP, von Elm E, Altman DG, Gotsche PC, Mulrow CD, Pocock SJ, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg* 2014;12(12):1500-24.
<http://dx.doi.org/10.1016/j.ijsu.2014.07.014>
142. 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.
143. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;158(9):915-20.
<http://dx.doi.org/10.1093/aje/kwg231>
144. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep* 2015;2(4):221-8.
<http://dx.doi.org/10.1007/s40471-015-0053-5>
145. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance, European Medicines Agency. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on methodological standards in pharmacoepidemiology (revision 8, 2020). EMA/95098/2010. Amsterdam: ENCePP; 2010.
http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml
146. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance, European Medicines Agency. ENCePP Checklist for study protocols (Revision 4). Amsterdam: ENCePP; 2018.
http://www.encepp.eu/standards_and_guidances/checkListProtocol.s.shtml
147. Strom BL, Kimmel SE, Hennessy S. *Pharmacoepidemiology*. 6th ed. Hoboken: Wiley-Blackwell; 2019.
148. Rothman K, Greenland S, Lash T. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams and Wilkins; 2012.
149. Eucomed. Eucomed medical device registries six key principles for the efficient and sustainable funding & reimbursement of medical devices. Brussels: EUcomed.
<https://www.fdanews.com/ext/resources/files/07/07-14-Eucomed-FR.pdf>
150. International Society for Pharmacoepidemiology. Guidelines for Good Pharmacoepidemiology Practices (GPP). Bethesda: ISPE; 2015.
<https://www.pharmacoepi.org/resources/policies/guidelines-08027/#Top>
151. Acker MA, Pagani FD, Stough WG, Mann DL, Jessup M, Kormos R, *et al.* Statement regarding the pre and post market assessment of durable, implantable ventricular assist devices in the United States. *Circ Heart Fail* 2013;6(1):e1-e11.
<http://dx.doi.org/10.1161/HHF.0b013e318279f6b5>
152. Reams BD, O'Malley CD, Critchlow CW, Lauffenburger JC, Brookhart MA. Changing patterns of use of osteoporosis medications in the years after launch: implications for comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2014;23(3):251-60.
<http://dx.doi.org/10.1002/pds.3545>
153. Ankarfeldt MZ, Thorsted BL, Groenwold RH, Adalsteinsson E, Ali MS, Klungel OH. Assessment of channeling bias among initiators of glucose-lowering drugs: A UK cohort study. *Clin Epidemiol* 2017;9:19-30.
<http://dx.doi.org/10.2147/CLEP.S124054>
154. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009;20(4):512-22.
<http://dx.doi.org/10.1097/EDE.0b013e3181a663cc>
155. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11(5):550-60.
<http://dx.doi.org/10.1097/00001648-200009000-00011>
156. You SC, Jung S, Swerdel JN, Ryan PB, Schuemie MJ, Suchard MA, *et al.* Comparison of first-line dual combination treatments in hypertension: real-world evidence from multinational heterogeneous cohorts. *Korean Circ J* 2020;50(1):52-68.
<http://dx.doi.org/10.4070/kcj.2019.0173>
157. Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big

- data: motivation-challenges-success factors. *Onkologie* 2018;24(Suppl 2):91-8.
<http://dx.doi.org/10.1007/s00761-018-0358-3>
158. Agency for Healthcare Research and Quality, Velentgas P, Dreyer N, Nourjah P, Smith S, Torchia M. Developing a protocol for observational comparative effectiveness research: a user's guide. Rockville: AHRQ; 2013.
https://effectivehealthcare.ahrq.gov/sites/default/files/related_files/user-guide-observational-cer-130113.pdf
159. Food and Drug Administration. Real-world evidence program. Framework for FDA's. Silver Spring: FDA; 2018.
<https://www.fda.gov/media/120060/download>
160. European Medicines Agency. STAMP Commission Expert Group. Brussels: EC; 2016.
https://ec.europa.eu/health/sites/health/files/files/committee/stamp/2016-03_stamp4/4_real_world_evidence_background_paper.pdf
161. Acker MA, Pagani FD, Stough WG, Mann DL, Jessup M, Kormos R, *et al.* Statement regarding the pre and post market assessment of durable, implantable ventricular assist devices in the United States: executive summary. *Circ Heart Fail* 2013;6(1):145-50.
<http://dx.doi.org/10.1161/HHF.0b013e318279f55d>
162. 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.
163. 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.
164. Torgerson DJ, Sibbald B. Understanding controlled trials. What is a patient preference trial? *BMJ* 1998;316(7128):360.
165. Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, *et al.* Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev Epidemiol Santé Publ* 2017;65 Suppl 4:S149-67.
<http://dx.doi.org/10.1016/j.respe.2017.05.004>
166. Daniel C, Serre P, Orlova N, Bréant S, Paris N, Griffon N. Initializing a hospital-wide data quality program. The AP-HP experience. *Comput Methods Programs Biomed* 2019;181:104804.
<http://dx.doi.org/10.1016/j.cmpb.2018.10.016>
167. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12(10):e1001885.
<http://dx.doi.org/10.1371/journal.pmed.1001885>
168. Conseil national de l'industrie. Programme « intelligence artificielle et santé » du comité stratégique de filière industries et technologies de santé. Recommandations du groupe de travail « usage des données en vie réelle dans la prise de décision ». Paris: ARIIS.
169. Cucherat M, Laporte S, Delaitre O, Behier JM, d'Andon A, Binlich F, *et al.* Des études mono-bras aux études de comparaison externe. Considérations méthodologiques et recommandations [Pre-proof]. *Thérapie* 2020;75(1):13-9.
<http://dx.doi.org/10.1016/j.therap.2019.12.002>
170. Haute Autorité de Santé. Les comparaisons indirectes. Méthode et validité. Saint-Denis La Plaine: HAS; 2009.
https://www.has-sante.fr/upload/docs/application/pdf/2010-11/rapport_de_synthese_-_comparaisons_indirectes_methodes_et_validite_-_juillet_2009.pdf
171. European Network for Health Technology Assessment. Comparators and comparisons: direct and indirect comparisons. Guideline. Brussels: EUnetHTA; 2015.
https://eunetha.eu/wp-content/uploads/2018/01/Comparators-Comparisons-Direct-and-indirect-comparisons_Amended-JA1-Guideline_Final-Nov-2015.pdf
172. Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, *et al.* Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 2012;15(6):940-7.
<http://dx.doi.org/10.1016/j.jval.2012.05.004>
173. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Decis Making* 2018;38(2):200-11.
<http://dx.doi.org/10.1177/0272989X17725740>
174. Food and Drug Administration. Rare diseases: natural history studies for drug development guidance for industry draft guidance. Silver Spring: FDA; 2019.
<https://www.fda.gov/media/122425/download>

175. Rosenbaum PR, Rubin D. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984;79(387):516-24.
176. Rosenbaum PR, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41-55.
177. 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.
178. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006;163(12):1149-56.
<http://dx.doi.org/10.1093/aje/kwj149>
179. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, *et al.* Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 2006;163(3):262-70.
<http://dx.doi.org/10.1093/aje/kwj047>
180. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46(3):399-424.
<http://dx.doi.org/10.1080/00273171.2011.568786>
181. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015;34(28):3661-79.
<http://dx.doi.org/10.1002/sim.6607>
182. Allan V, Ramagopalan SV, Mardekian J, Jenkins A, Li X, Pan X, *et al.* Propensity score matching and inverse probability of treatment weighting to address confounding by indication in comparative effectiveness research of oral anticoagulants. *J Comp Eff Res* 2020;9(9):603-14.
<http://dx.doi.org/10.2217/cer-2020-0013>
183. Rosenbaum PR. Discussing hidden bias in observational studies. *Ann Intern Med* 1991;115(11):901-5.
184. 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.
185. Maura G, Blotière PO, Bouillon K, Billionnet C, Ricordeau P, Alla F, *et al.* Comparison of the short-term risk of bleeding and arterial thromboembolic events in nonvalvular atrial fibrillation patients newly treated with dabigatran or rivaroxaban versus vitamin K antagonists: a French nationwide propensity-matched cohort study. *Circulation* 2015;132(13):1252-60.
<http://dx.doi.org/10.1161/circulationaha.115.015710>
186. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;29(4):722-9.
187. Chen Y, Briesacher B. Use of instrumental variable in prescription drug research with observational data : a systematic review. *J Clin Epidemiol* 2011;64(6):687-700.
188. 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.
189. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010;19(6):537-54.
190. 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.
191. Baser O. Too much ado about instrumental variable approach: is the cure worse than the disease? *Value Health* 2009;12(8):1201-9.

Abréviations et acronymes

ANSM	Agence nationale de sécurité du médicament et des produits de santé
CNAMTS	Caisse nationale de l'assurance maladie des travailleurs salariés
CNEDiMITS	Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé
CSS	Code de la sécurité sociale
DGOS	Direction générale de l'Offre des soins
DGS	Direction générale de la Santé
DM	Dispositif médical
DMC	Dispositif médical connecté
DSS	Direction de la Sécurité sociale
ECR	Essai contrôlé randomisé
EMA	<i>European Medicines Agency</i>
FDA	<i>Food and Drug Administration</i>
FEDEPSAD	Fédération des prestataires de santé à domicile
HAS	Haute Autorité de santé
HTA	<i>Health Technology Assessment</i>
IC	Intervalle de confiance
INAHTA	<i>International Network of Agencies for Health Technology Assessment</i>
INCa	Institut national du cancer
ITT	Intention de traiter
LPPR	Liste des produits et prestations remboursables
MOST	<i>Multiphase Optimisation Strategy</i>
NHS	<i>National Health Service</i>
NICE	<i>National Institute for Health and Care Excellence</i>
OMS	Organisation mondiale de la santé
PP	Per-protocole
PREMs	<i>Patient Reported Experience Measures</i>
PROMs	<i>Patient Reported Outcome Measures</i>
SED	Service évaluation des dispositifs (de la HAS)
SMART	<i>Sequential Multiple Assignment Randomised Trials</i>
SNDS	Système national des données de santé
YHEC	<i>York Health Economic Consortium</i>

Glossaire

Big data : correspond à des bases de données de taille très importante (de l'ordre de 100 millions de lignes et colonnes, voire plus).

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 influent sur 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 lorsque 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 au défaut d'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 bêta : 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 bêta 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 dû au hasard mais en rapport avec le traitement administré).

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

Retrouvez tous nos travaux sur

www.has-sante.fr

