

**Compte-rendu des  
2<sup>ème</sup> rencontres  
méthodologiques  
de l'AFRETH**

**Judi 25 Octobre 2012**

**Paris, maison de l'Amérique Latine**

**Innovations  
méthodologiques dans  
l'évaluation des  
interventions complexes**

# **2èmes rencontres méthodologiques de l'AFRETH**

**Jeudi 25 Octobre 2012**

**Paris, Maison de l'Amérique Latine**

## **Innovations méthodologiques dans l'évaluation des interventions complexes**

**Président** : Pr D. Costagliola (INSERM, Université Pierre et Marie Curie, Paris)

**Rapporteur** : Dr A. Dechartres (Paris)

**Participants** : Dr P. Blin, Pr D. Costagliola, Dr A. Dechartres, Pr C. Funck-Brentano, Pr G. Chatellier, Pr J. Massol, Dr R. Forestier, Dr A. Françon, Dr R.C. Mancret, Pr P. Ravaud, Dr L Riou-Franca, Pr C-F Roques, Mr W. Tabone, Pr D. Torgerson

### **Programme**

#### **Matin (9.30-12.30)**

**\* Innovative designs for the assessment of complex interventions.**

D. Torgerson (York University, UK)

**\* Place respective de l'analyse des données observationnelles et des essais randomisés dans l'évaluation thérapeutique des interventions complexes.** Ph. Ravaud (CEC, Paris)

**\* Analyse des travaux menés sous l'égide de l'AFRETH, difficultés rencontrée.** C-F. Roques (Afreth, Paris)

**\* Points importants et particularités méthodologiques dans l'évaluation des interventions thérapeutiques complexes.** Ph. Ravaud (CEC, Paris)

#### **Après midi (14h-17h)**

**Transposition à la problématique de l'évaluation des cures thermales – conséquences pratiques**

**Synthèse et conclusions par la Présidente**

## Sommaire

Schémas expérimentaux innovants pour l'évaluation des interventions complexes.....	5
Introduction.....	5
« Balanced incomplete block ».....	5
Principe .....	5
Avantages.....	6
Inconvénients.....	6
Schéma de Zelen .....	6
Principe .....	6
Avantages.....	7
Inconvénients.....	7
« Patient preference design » ou schéma de Brewin et Bradley .....	7
Rationnel.....	7
Principe .....	8
Alternative « Fully randomised preference design » .....	8
Schéma « Cohort multiple randomised controlled trial » .....	9
Principe .....	9
Avantages.....	9
Inconvénients.....	10
Analyse .....	10
Place respective de l'analyse des données observationnelles et des essais randomisés dans l'évaluation thérapeutique des interventions complexes .....	13
Comparaison des études observationnelles et des essais contrôlés randomisés .....	13
Avantages et inconvénients des essais contrôlés randomisés et des études observationnelles .....	13
Méthodologie des études observationnelles .....	13
Problématiques actuelles des études observationnelles.....	14
Conclusion et perspectives.....	15
Analyse des travaux menés sous l'égide de l'AFRETH : difficultés rencontrées.....	16
Projets conduits sous l'égide de l'AFRETH.....	16
Méthodologie des ECRs conduits.....	17
Recrutement.....	17
Résultats des ECRs conduits sous l'égide de l'AFRETH.....	18
Conclusion.....	18

Points importants et particularités méthodologiques dans l'évaluation des interventions complexes .....	19
Définitions et typologie .....	19
Particularités méthodologiques de l'évaluation des traitements non-pharmacologiques ....	20
Aveugle .....	20
Particularités liées au caractère complexe de l'intervention .....	21
Conséquences pratiques pour l'évaluation thérapeutique des cures thermales .....	25
Réflexion sur les critères de jugement .....	25
Quels types de critère de jugement .....	25
Horizon temporel .....	25
Evaluation des évènements indésirables .....	26
Réflexion sur ce qu'on souhaite évaluer .....	26
Quelle intervention ? .....	26
Quelle durée ? .....	26
Standardisation de l'intervention .....	26
Recrutement des patients .....	27
Difficultés de recrutement .....	27
Problème de représentativité des patients .....	27
Sur quelles bases sélectionne-t-on des patients atteints d'une certaine pathologie?.....	27
Utilisation des nouveaux outils technologiques .....	28
Utilisation des bases de données médico-administratives.....	28
Propositions d'amélioration dans le cadre de la mise en place de projets visant à évaluer les cures thermales .....	29
Références .....	30

# Schémas expérimentaux innovants pour l'évaluation des interventions complexes

## *Introduction*

Les essais randomisés contrôlés constituent la référence en évaluation thérapeutique pour limiter les biais et les facteurs de confusion. Un biais est défini comme une erreur systématique dans l'évaluation d'un paramètre. Un facteur de confusion est défini comme un facteur ayant un lien à la fois avec le facteur d'exposition étudié et la maladie étudiée sans faire parti de la chaîne causale liant ces 2 éléments. La randomisation permet d'obtenir des groupes comparables pour tous les facteurs pronostiques connus et inconnus. D'autres facteurs peuvent introduire des biais après la randomisation. Il s'agit notamment de l'attrition des patients particulièrement si elle diffère selon le bras, des cross-overs (le fait qu'un patient alloué à une certaine intervention va finalement recevoir celle de l'autre bras) ce qui risque de sous estimer l'effet traitement, de l'évaluation des critères de jugement en ouvert qui peut entraîner un biais d'évaluation, de la préférence des patients pour une intervention particulière qui peut également introduire des biais dans l'évaluation et de l'effet Hawthorne c'est-à-dire le fait que l'attitude du patient peut être influencé par la présence de l'investigateur. Quand l'intervention est un médicament, il est souvent relativement simple d'utiliser comme comparateur soit un placebo soit un traitement actif d'apparence similaire au traitement expérimental ce qui va permettre de limiter les biais d'évaluation, les effets liés à la préférence pour l'une des interventions et les cross-overs. Pour les interventions non-médicamenteuses, il est souvent difficile voire impossible d'utiliser comme comparateur soit un placebo ou soit une fausse intervention (sham intervention) ce qui peut entraîner des biais post-randomisation.

Certains schémas expérimentaux permettent de limiter ces biais.

## **« *Balanced incomplete block* »**

### *Principe*

Ce schéma expérimental peut être utilisé pour gérer les problèmes liés à l'effet Hawthorne et va permettre de limiter l'attrition liée à une déception pour les patients du groupe contrôle. Ce schéma est fréquemment utilisé pour évaluer des recommandations de pratique clinique ou des formations. Prenons l'exemple d'un essai dans lequel on veut évaluer une formation destinée aux thérapeutes portant sur la prise en charge de patients lombalgiques chroniques. Si on fait un essai classique à 2 bras parallèles, les thérapeutes randomisés dans le bras contrôle peuvent être déçus. Dans le bras intervention, le fait d'avoir la formation peut avoir un effet quelque soit le contenu de la formation. Un schéma expérimental de type « *Balanced incomplete block* » serait dans cette situation de randomiser les thérapeutes entre une formation pour prendre en charge des patients lombalgiques

et une formation pour prendre en charge des patients avec polyarthrite rhumatoïde. Les thérapeutes de chaque bras prendront en charge à la fois des patients lombalgiques chroniques et des patients avec une polyarthrite rhumatoïde. Pour évaluer l'efficacité de la formation sur la prise en charge de patients lombalgiques chroniques, le critère de jugement sera comparé entre les patients lombalgiques chroniques ayant ou non des thérapeutes formés. Une comparaison similaire sera faite pour la polyarthrite rhumatoïde. Ce schéma expérimental permet ainsi d'évaluer 2 interventions différentes.

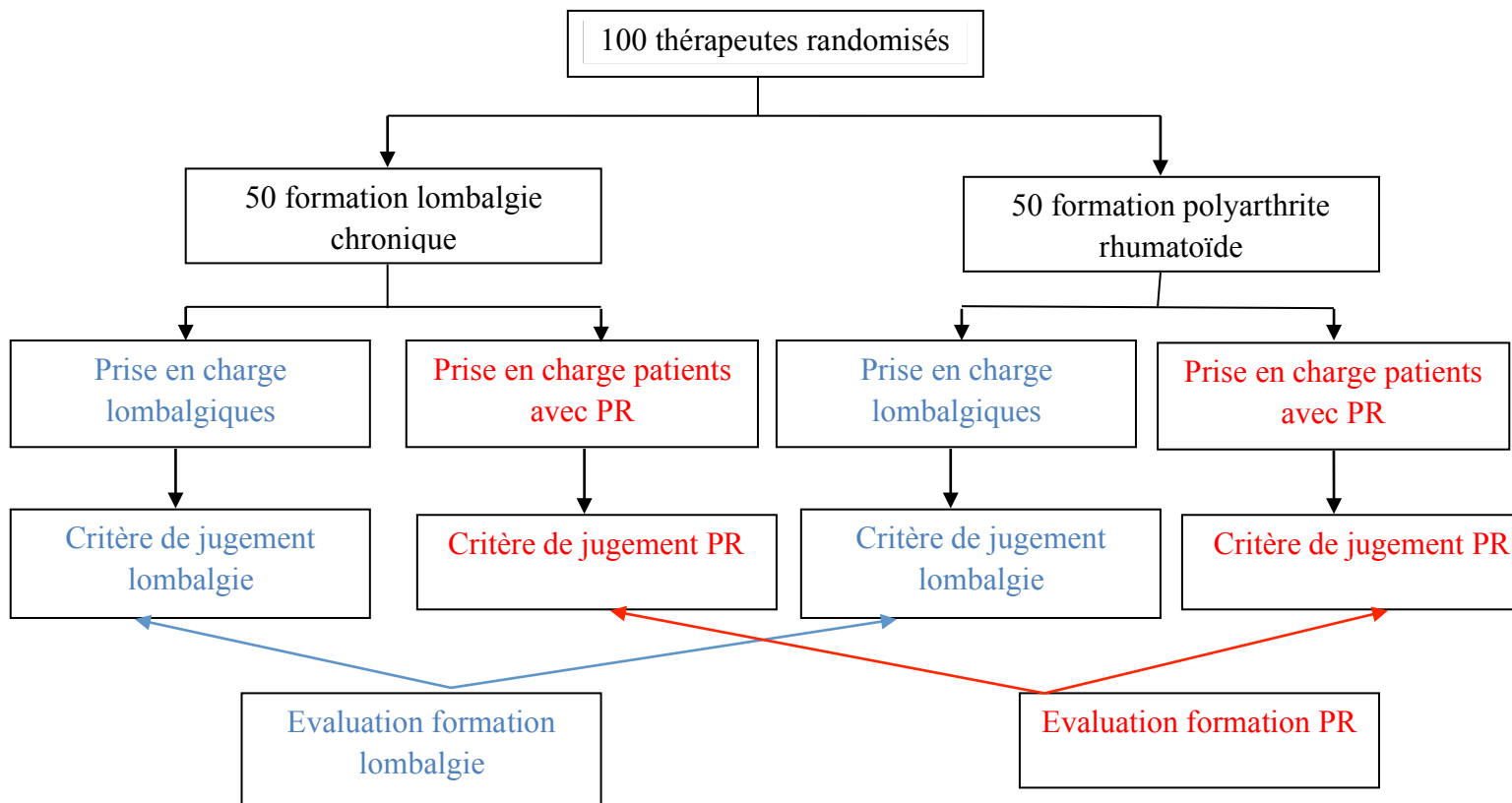


Figure 1: Schéma expérimental de type « Balanced incomplete block »

### Avantages

Ce schéma expérimental permet de prendre en compte l'effet Hawthorne ainsi que la déception de ne pas être dans le bras intervention.

### Inconvénients

Les situations où on peut l'utiliser sont assez limitées. Ce schéma nécessite de recruter avant la randomisation sans quoi le recrutement peut être biaisé.

## Schéma de Zelen

### Principe

Proposé par Marvin Zelen, il s'agit d'un schéma où les patients sont randomisés avant l'obtention du consentement et celui-ci n'est demandé que pour les patients randomisés dans le bras expérimental(1). Les patients randomisés dans le bras

contrôle (prise en charge habituelle) ne sont informés de l'essai. Ce schéma bien connu est finalement relativement peu utilisé(2, 3), parfois dans les campagnes de dépistage.

### Avantages

Ce schéma permet d'éviter l'effet de la préférence des patients en évitant la déception de ne pas être dans le bras intervention mais aussi l'effet Hawthorne.

### Inconvénients

Ce schéma pose des problèmes éthiques. De plus, si beaucoup de patients refusent de donner leur consentement dans le bras intervention cela risque de sous-estimer l'effet traitement. Si les patients apprennent qu'ils sont dans le groupe contrôle, leur déception risque d'être majorée.

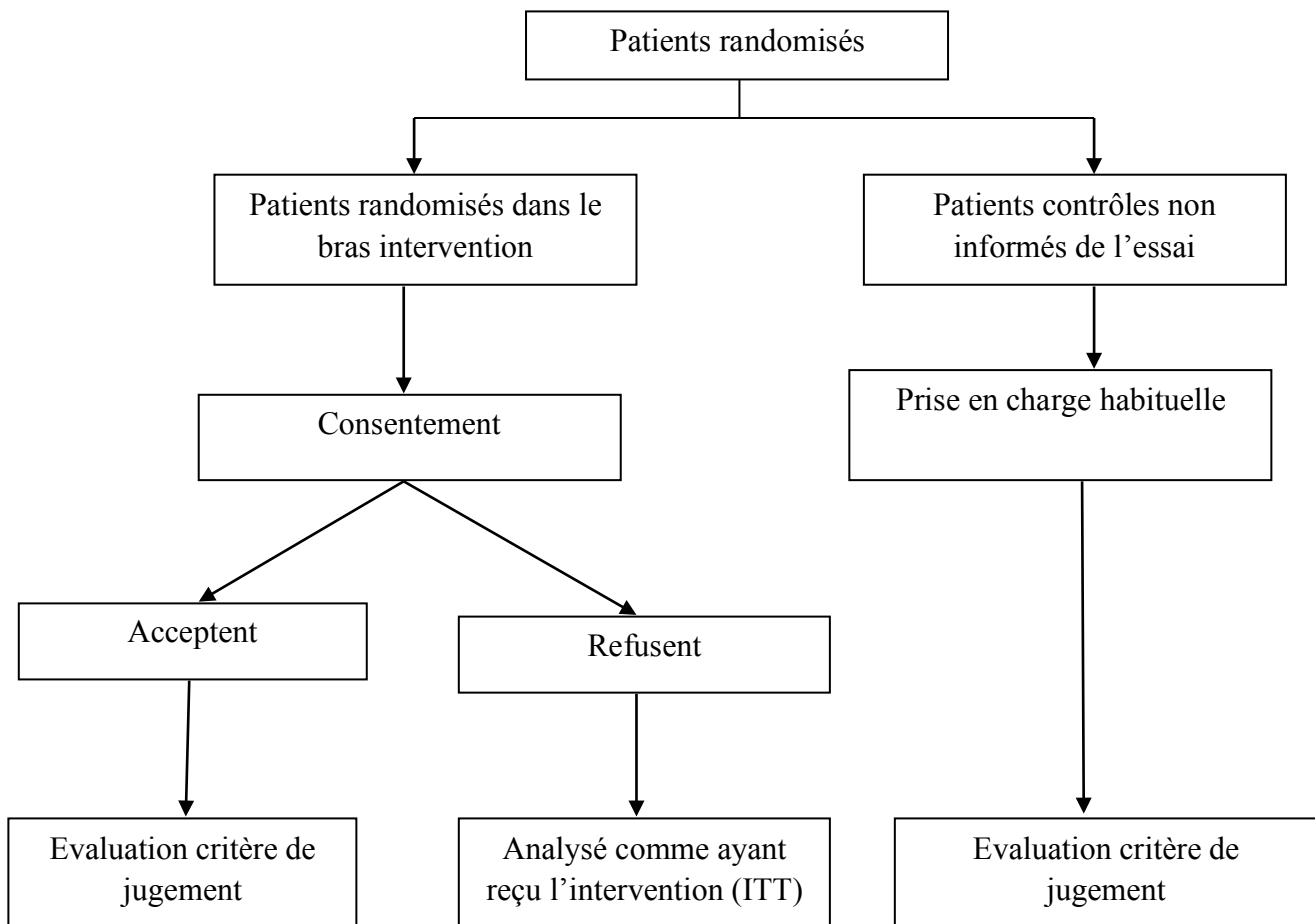


Figure 2: Schéma expérimental de Zelen

## « Patient preference design » ou schéma de Brewin et Bradley

Rationnel

Les patients ont très fréquemment une préférence pour l'une des interventions qui leur sont offertes. Cela pose un problème quand on fait un essai randomisé classique car ceux qui ont une préférence pour le traitement expérimental et sont randomisés dans le bras contrôle risquent d'être déçus et ceux qui sont randomisés dans le bras expérimental risquent d'avoir des attentes démesurées ce qui va biaiser l'estimation de l'effet traitement.

*Principe*

Les patients sont interrogés sur leur préférence ; ceux qui en ont une reçoivent l'intervention qu'ils préfèrent tandis que ceux qui n'en ont pas sont randomisés(4).

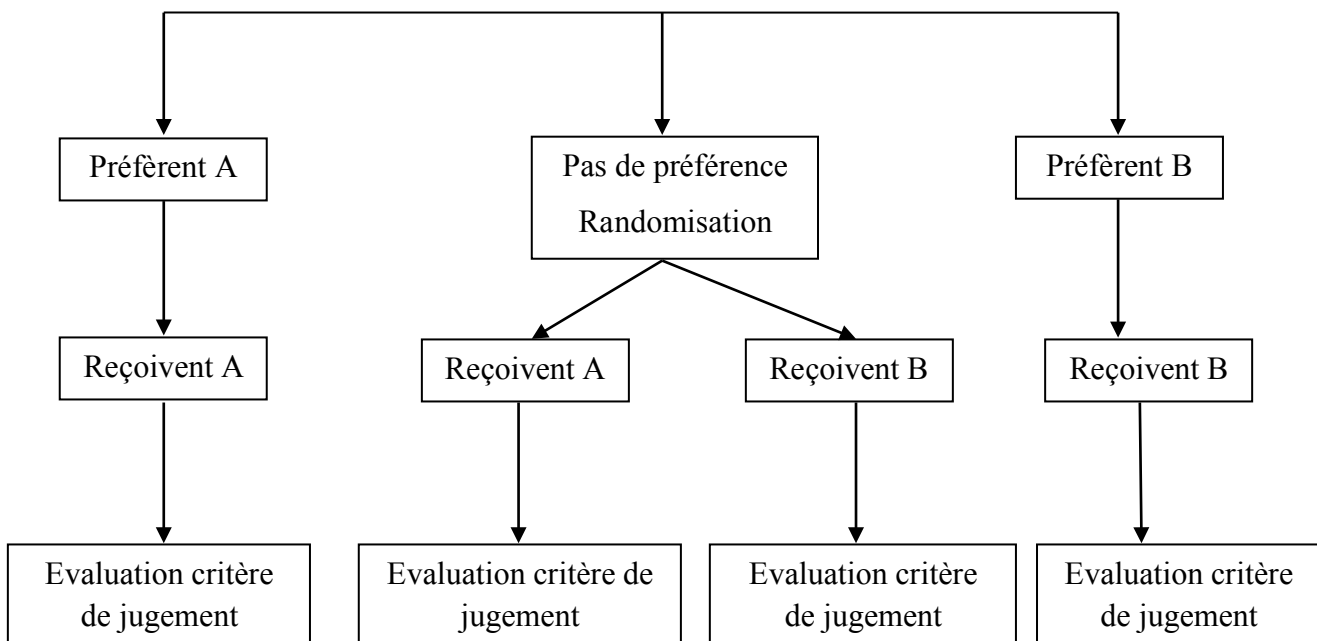


Figure 3: « Preference design »

Ce schéma pose des problèmes d'interprétation. En effet, les résultats sont difficilement interprétables pour les patients ayant une préférence pour l'une ou l'autre des interventions. La comparaison sur les patients randomisés peut porter sur un nombre très limité de patients.

*Alternative « Fully randomised preference design »*

Tous les patients sont interrogés sur leur préférence quant au traitement et tous sont randomisés quelque soit leur préférence. L'ensemble des patients ayant donné leur consentement sont ainsi pris en compte. L'interaction avec la préférence est évaluée mais en pratique la puissance pour l'évaluer est très faible.



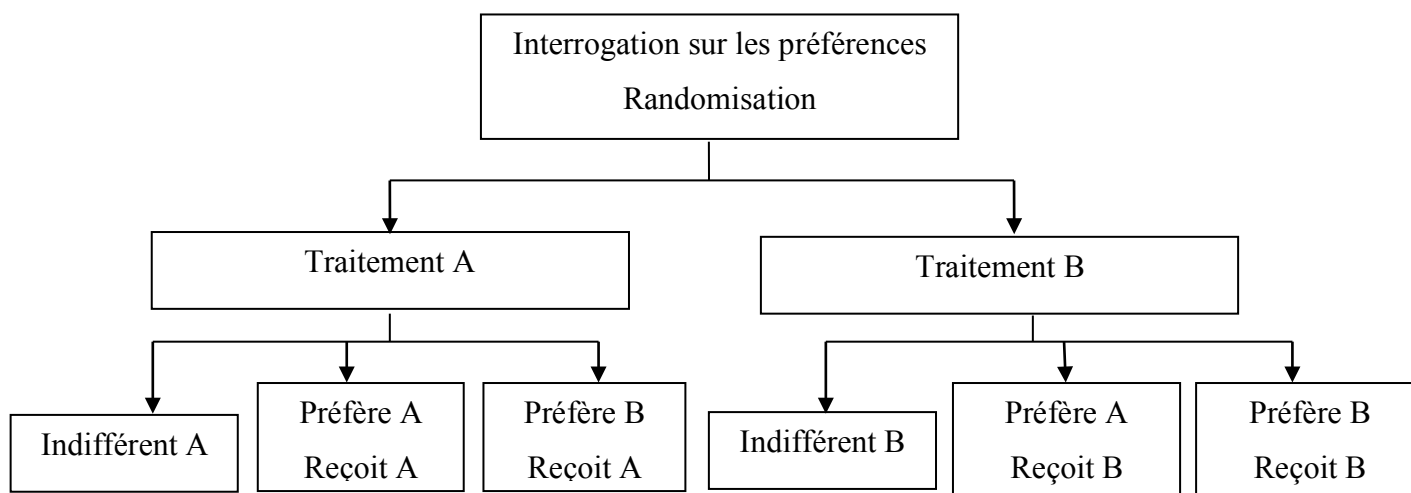


Figure 4: Schéma « Fully randomised preference design »

### **Schéma « Cohort multiple randomised controlled trial »**

#### *Principe*

Le principe est de tirer au sort des patients à qui va être proposée une intervention au sein d'une large cohorte pré-existante. Lors de la constitution de la cohorte, les patients sont interrogés sur les traitements qu'ils pourraient envisager. Les patients signent un consentement dans le cadre de la cohorte concernant la collecte de données les concernant. Un consentement à recevoir une intervention particulière n'est demandé qu'aux patients qui ont été tirés au sort pour recevoir cette intervention. Les patients ne sont pas informés des interventions qui ne leur sont pas proposées. Les patients ne sont pas informés des interventions qu'ils ne reçoivent pas. Ils ne sont pas non plus informés que les interventions sont attribuées de manière aléatoire. Une certaine période doit s'écouler avant la réalisation de l'essai pour s'assurer que les patients vont répondre(5).

#### *Avantages*

Ce schéma présente plusieurs avantages. Tout d'abord, le taux d'attrition après tirage au sort est faible. En effet, en raison du délai entre la constitution de la cohorte et le tirage au sort, les sorties d'étude sont censées avoir lieu avant le tirage au sort. Ensuite, les patients ne sont informés que des interventions qui leur sont proposées ce qui permet de limiter une éventuelle déception des patients lorsqu'il existe une préférence pour une intervention particulière. Enfin, ce schéma est probablement plus pragmatique(6), plus représentatif de la vraie vie qu'un essai

randomisé classique. Si la cohorte est suffisamment grande, elle permettra également de comparer plusieurs stratégies thérapeutiques. La validité externe de l'étude peut être évaluée grâce au suivi des patients non tirés au sort dans le cadre de la cohorte. En revanche, ceux qui n'ont pas renvoyé le questionnaire initialement ne sont pas suivis ce qui limite la validité externe mais c'est un problème commun à toutes les études.

### *Inconvénients*

Ce schéma ne peut s'utiliser que dans le cadre de maladies chroniques, il n'est pas faisable pour évaluer des stratégies thérapeutiques dans le cadre de pathologies incidentes (par exemple, 2 stratégies thérapeutiques pour la prise en charge d'une fracture). Les cross-over entre les bras peuvent également être plus importants que pour un essai randomisé classique ce qui risque de diluer l'effet traitement.

### *Analyse*

L'analyse devrait reposer sur le principe de l'intention de traiter. Néanmoins, en cas de cross-over important, il est possible d'utiliser une autre stratégie d'analyse « Complier average causal effect »(7) qui va permettre une estimation non-biaisée de l'effet traitement pour les participants compliants avec l'intervention. Pour cela, il faut que 2 conditions soient réunies :

- 1) La proportion de patients dans le groupe contrôle qui auraient refusé le traitement si on leur avait proposé est similaire à la proportion de patients refusant le traitement dans le bras intervention.
- 2) Le fait de proposer le traitement n'a pas d'effet sur le critère de jugement

En pratique, la première condition est toujours vraie car il y a tirage au sort. La seconde hypothèse est probablement vraie dans beaucoup de circonstances. De plus, cette hypothèse est probablement plus défendable scientifiquement que celle sur laquelle repose l'analyse per protocol (c'est-à-dire que la population compliante ne diffère pas de la population randomisée).

En pratique, le fait d'interroger à l'avance les patients sur les traitements qu'ils pourraient considérer permet de diminuer les cross-over.

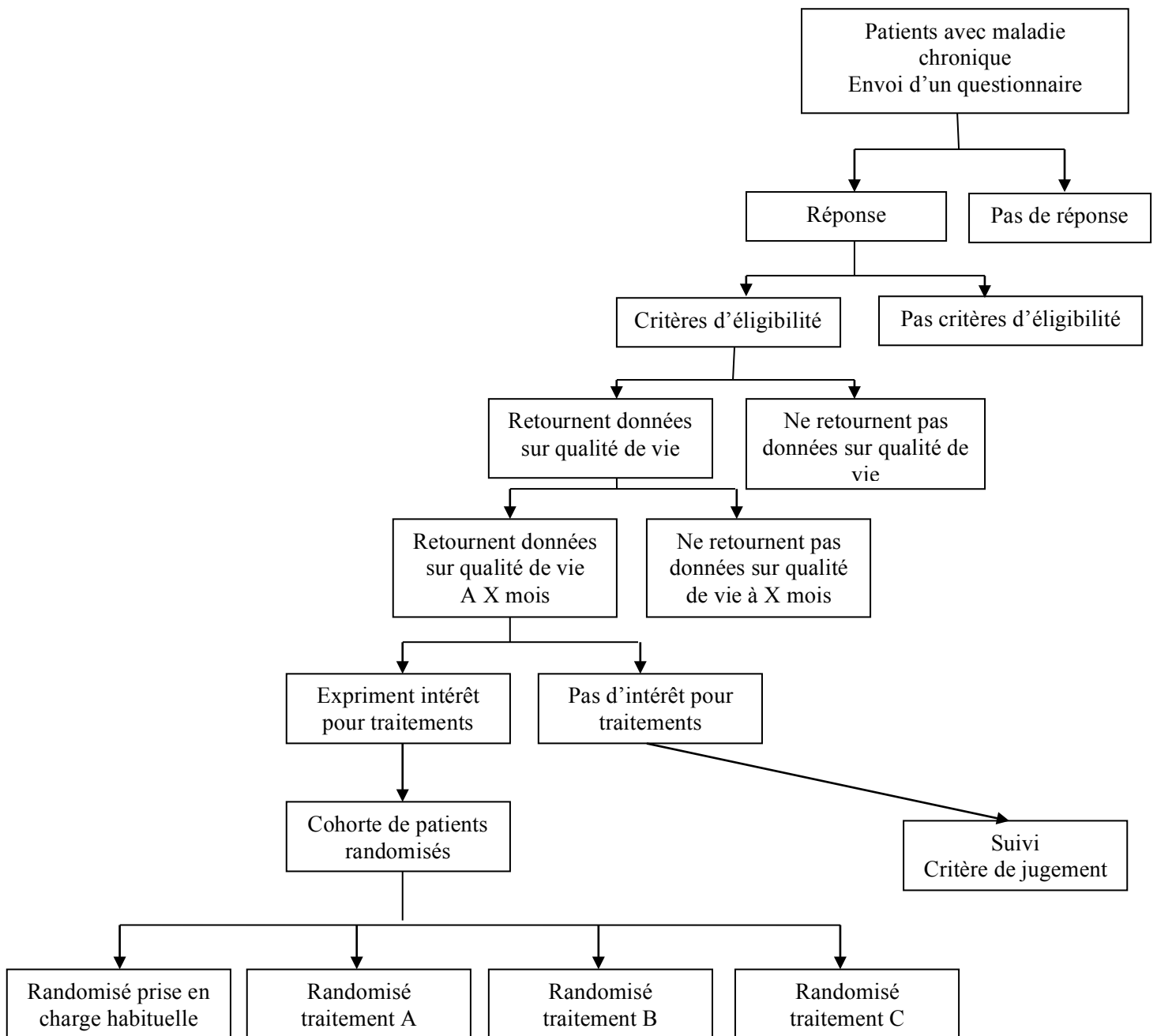


Figure 5: Schéma du « cohort multiple randomised controlled trial »

**Tableau 1 : Résumé des types d'étude**

Schéma expérimentaux	RCT	Balanced incomplete block	Zelen	Preference trial	Cohorte randomisée
<b>Principe</b>	Sélection de patients qui donnent ensuite leur consentement pour être randomisés pour recevoir un traitement expérimental ou contrôle	<ul style="list-style-type: none"> <li>- Utilisé pour évaluer des recommandations de pratique clinique ou des formations</li> <li>- Randomisation pour recevoir formations dans 2 pathologies différentes</li> <li>- Dans chaque bras, prise en charge patients avec l'une ou l'autre des pathologies</li> <li>- Comparaison des résultats entre thérapeutes formés et non formés</li> </ul>	Randomisation précède le consentement et seuls les patients qui reçoivent le traitement expérimental doivent donner leur consentement	2 schémas : <ul style="list-style-type: none"> <li>- Soit prise en compte de la préférence : randomisation et comparaison que chez ceux n'ayant pas de préférence</li> <li>- Soit randomisation qq soit la préférence et évaluation interaction préférence et effet traitement</li> </ul>	Recrutement d'une large cohorte Identification de patients avec suivi et intéressés par tel ou tel traitement Randomisation et consentement des patients recevant l'intervention
<b>Avantages</b>	Contrôle des biais	<ul style="list-style-type: none"> <li>- Limite la déception dans le groupe contrôle</li> <li>- Limite l'effet Hawthorne</li> </ul>	<ul style="list-style-type: none"> <li>- Limite la déception dans le groupe contrôle</li> <li>- Limite l'effet Hawthorne</li> </ul>	Prise en compte de la préférence	<ul style="list-style-type: none"> <li>- Attrition faible</li> <li>- Limite la déception dans le groupe contrôle</li> <li>- Pragmatique</li> <li>- Permet de comparer plusieurs traitements</li> </ul>
<b>Inconvénients</b>	<ul style="list-style-type: none"> <li>- Mauvaise validité externe : patients non représentatifs</li> <li>- Difficultés de recrutement (refus de la randomisation)</li> <li>- Si essai en ouvert, préférence des patients et risque de déception dans le groupe contrôle</li> </ul>	<ul style="list-style-type: none"> <li>- Application limitée</li> <li>- Nécessite de recruter avant randomisation sinon risque de biais de recrutement</li> </ul>	<ul style="list-style-type: none"> <li>- Problème éthique (information incomplète)</li> <li>- Dilution de l'effet traitement si beaucoup de patients refusent l'intervention</li> </ul>	<ul style="list-style-type: none"> <li>- 1<sup>er</sup> schéma : faible nombre de patients randomisés (faible puissance)</li> <li>- 2<sup>nd</sup> schéma : faible puissance statistique pour mettre en évidence une interaction</li> </ul>	<ul style="list-style-type: none"> <li>- Limité aux pathologies chroniques</li> <li>- Dilution de l'effet traitement si cross-over</li> </ul>

# **Place respective de l'analyse des données observationnelles et des essais randomisés dans l'évaluation thérapeutique des interventions complexes**

## ***Comparaison des études observationnelles et des essais contrôlés randomisés***

Pendant de nombreuses années, l'essai contrôlé randomisé a occupé une place prépondérante et indiscutable dans le domaine de l'évaluation thérapeutique. Les études observationnelles ont longtemps été considérées comme inadaptées en raison de leurs biais importants(8). Quelques études ont comparé les résultats des essais randomisés contrôlés et des études observationnelles(9-12). Par exemple, l'étude de Ioannidis publiée dans le JAMA en 2001(12) montre qu'il n'y a pas de surestimation systématique de l'effet traitement. Globalement ces études montrent une bonne corrélation entre les résultats des études observationnelles et des essais contrôlés randomisés réalisés sur le même sujet. Les résultats retrouvés dans les études observationnelles sont donc le plus souvent comparables aux résultats des essais mais pas toujours. Il existe des exemples tels que l'hormonothérapie substitutive après la ménopause pour lesquels résultats d'études observationnelles et d'essais randomisés divergent(13, 14).

## ***Avantages et inconvénients des essais contrôlés randomisés et des études observationnelles***

Actuellement, la communauté scientifique est consciente des forces des essais contrôlés randomisés notamment par le contrôle des biais mais également de ses faiblesses notamment le manque de validité externe. En effet, les malades recrutés dans les essais contrôlés randomisés sont très différents de ceux traités dans la vraie vie. Il est actuellement reconnu que les études observationnelles ont des avantages que les essais randomisés n'ont pas : elles sont plus proches de la vie réelle car elles incluent des patients qui sont généralement exclus des essais randomisés, elles permettent également un suivi plus long que dans les essais randomisés(15, 16).

## ***Méthodologie des études observationnelles***

Les principaux types d'études observationnelles sont les études de cohorte et les registres qui peuvent concerner une maladie particulière ou un traitement particulier. La mode est également d'utiliser de grosses bases de données administratives et éventuellement de les regrouper. L'intérêt actuel pour l'utilisation d'études observationnelles en évaluation thérapeutique provient du fait que des techniques

statistiques ont été développées afin de prendre en compte les biais et facteurs de confusion. La principale technique utilisée est le score de propension qui permet de mimer un essai randomisé contrôlé(17). En effet, une étude expérimentale a montré que le résultat d'une comparaison avec score de propension était très proche du résultat d'un essai randomisé. Le score de propension peut être défini comme la probabilité pour un patient donné de recevoir le traitement conditionnellement à ses caractéristiques(18). Le score de propension dans une étude peut être observé à partir des données. Il permet d'équilibrer les covariables mesurées entre les groupes traités et non traités. L'utilisation du score de propension repose sur 3 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 observées
- 3) Indépendance conditionnelle de l'attribution du traitement et des critères de jugement

Il y a un véritable engouement pour le score de propension et une augmentation très importante de son utilisation en médecine. Il existe différentes méthodes pour prendre en compte le score de propension dans l'analyse : l'ajustement, l'appariement, la stratification, la pondération inverse. Ces différentes méthodes d'analyse répondent à des questions différentes(19).

Néanmoins, le score de propension ne constitue pas une méthode « magique » qui permet de s'affranchir de tous les biais. Il faudra vérifier systématiquement comment il a été construit et comment il est utilisé et s'il a bien pris en compte les principaux facteurs associés au critère de jugement et les principaux déterminants du traitement.

### ***Problématiques actuelles des études observationnelles***

La qualité des données recueillies constituent une problématique importante des études observationnelles. Certaines données importantes peuvent être manquantes dans les registres, les bases de données administratives ou lorsque les études s'appuient sur des données recueillies dans le passé. A titre d'exemple, les données concernant la consommation tabagique sont manquantes dans certaines bases de données administratives et ne peuvent par conséquent pas être prises en compte dans l'analyse. On peut s'interroger sur la pertinence des résultats observés dans cette situation et de manière plus générale sur la construction d'un score de propension à partir des données disponibles et non à partir de l'ensemble des données qui auraient été pertinentes. Ainsi, dans le domaine cardiovasculaire, est-il raisonnable d'avoir un score de propension sans le tabac. Une solution envisageable mais qui devrait être évaluée est le high dimensional propensity score qui consiste à construire un score de propension à partir d'un très grand nombre de variables. L'absence d'une variable importante ne serait alors pas un problème puisqu'elle serait prise en compte grâce au grand nombre de variables mesurées. Les études avec score de propension peuvent également poser des problèmes de validité

externe en fonction de la technique statistique utilisée notamment l'appariement. Il faudra donc vérifier le degré d' « overlap » entre les patients traités et non traités. La pondération inverse permet de prendre en compte l'ensemble des patients.

Par ailleurs, la qualité des données recueillies dans le cadre d'études observationnelles est peut être moins bonne que dans le cadre des essais randomisés contrôlés qui sont rigoureusement monitorés. Il est donc important de prévoir du personnel pour vérifier la qualité des données recueillies dans les études observationnelles en s'attachant notamment aux données manquantes et aberrantes. La mauvaise qualité des données n'entraînera un biais que si la mauvaise qualité des données est liée au traitement reçu et au critère de jugement.

En résumé, la qualité des données recueillies ainsi que l'exhaustivité des données pertinentes devra être systématiquement évaluée dans les études observationnelles.

### ***Conclusion et perspectives***

Les études observationnelles et les essais randomisés ne répondent pas aux mêmes questions et ces deux types d'étude sont nécessaires dans le domaine de l'évaluation thérapeutique(20). Il y a eu une amélioration considérable de la qualité des études observationnelles par transposition des critères de qualité d'un protocole des essais randomisés aux études observationnelles. Les études observationnelles peuvent permettre d'estimer l'ampleur de l'effet traitement dans la vraie vie.

## Analyse des travaux menés sous l'égide de l'AFRETH : difficultés rencontrées

La cure thermale est une intervention thérapeutique complexe dont la démonstration des bénéfices repose actuellement sur les essais randomisés contrôlés et les méta-analyses. Les investigations pré-cliniques ont porté sur l'évaluation des mécanismes d'action notamment le rôle des minéraux. Les bases historiques de l'évaluation reposent sur l'expérience des médecins et sur des cohortes de patients avec recueil des données à partir des dossiers médicaux. La littérature retrouve entre 100 et 120 essais contrôlés randomisés. Une méta-analyse a été conduite en 2006 incluant 46 essais randomisés contrôlés avec des critères de jugement cliniques. Une enquête d'opinion sur les cures thermales a été réalisée par TNS Healthcare auprès des patients en 2006. L'étude a porté sur 112419 questionnaires. Les résultats montraient que 71% des patients rapportaient une diminution de la douleur physique, 50% une diminution de la consommation de médicaments. L'étude de l'opinion des curistes sur les effets à 6 mois montrait que 97% d'entre eux rapportaient au moins un bénéfice : 74% rapportaient moins de douleur physique, 57% moins de consommation de médicaments. Quand on demandait aux patients si la cure thermale était plus efficace ou moins efficace pour soulager la douleur que les médicaments, 46% répondaient plus efficace, 49% aussi efficace et seulement 5% moins efficace. Quand on interrogeait les patients sur ce qu'ils pensaient être efficace dans la cure thermale, 47% répondaient que seule la cure thermale était efficace ; pour 60% des patients, l'environnement était également important.

### ***Projets conduits sous l'égide de l'AFRETH***

En 8 ans, l'AFRETH a lancé 7 appels à projet pour un montant global de 7 millions d'euros. Sur 101 pré-projets, 74 étaient éligibles, 41 ont été reconnus comme valable sur le plan scientifique et 26 ont été finalement financés.

	Avant-projets	Eligibles	Protocoles complets	Validité scientifique	Décision de financement
Médecine Thermale	90	68	68	35	23
Sécurité produits	4	3	3	3	2
Biologie	7	3	3	3	1
<b>Total</b>	<b>101</b>	<b>74</b>	<b>74</b>	<b>41</b>	<b>26</b>



Six essais contrôlés randomisés ont été conduits pour évaluer le bénéfice médical : trois ont été publiés (Stop-Tag portant sur l'anxiété généralisée(21), Thermarthrose sur l'arthrose du genou(22), Maathermes sur le surpoids et l'obésité(23)). Trois autres sont terminés. L'article rapportant les résultats est en lecture (étude Pacthe sur les suites de cancer du sein), en soumission (étude Rotatherm portant sur les tendinopathies de l'épaule), en achèvement de rédaction (étude Thermes&Veines sur l'insuffisance veineuse chronique).

Cinq études sont actuellement mises en place, certaines portent sur l'éducation thérapeutique. D'autres études sont en cours notamment un essai chez des patients avec bronchopneumopathie chronique obstructive, un autre chez des lombalgiques chroniques et un sur l'évaluation gériatrique.

### ***Méthodologie des ECRs conduits***

Les ECRs avaient généralement comme comparateur la prise en charge habituelle. Un essai a comparé le bénéfice de la cure thermale par rapport à un médicament anti-dépresseur.

Tous les ECRs avaient un calcul d'effectif basé sur le critère de jugement principal. L'analyse statistique finale était en intention de traiter pour tous les essais à l'exception d'un dont l'analyse a utilisé le score de propension en raison d'un grand nombre de cross-over entre les 2 bras. Le schéma de Zelen a été utilisé pour 2 essais (thermarthrose et Maathermes). Dans 2 essais, les interventions expérimentales et contrôles étaient respectivement traitement immédiat et traitement retardé.

Au total durant 6 ans, l'AFRETH a conduit 6 ECRs recrutant en tout 1818 patients soit en moyenne 303 patients par essai. En comparaison, la méta-analyse de 2006 sur 14 ans portant sur 46 essais a inclus 2764 patients soit en moyenne 61 par essai.

### ***Recrutement***

#### **Paradigmes de recrutement :**

- 1) Recrutement direct des patients par voie des médias notamment de la presse écrite  
Dans l'essai Thermarthrose, 630 patients ont répondu et 462 étaient éligibles
- 2) Recrutement via les médecins spécialistes  
Dépend des spécialistes. Ce mode de recrutement n'a pas bien marché pour les ORL libéraux dans l'essai Ectottit : 129 ont été contactés, 23 ont accepté de participer, 12 ont recruté un total de 26 patients sur les 160 prévus
- 3) Recrutement via les médecins généralistes  
Dans l'essai Maathermes, sur les 7000 médecins généralistes contactés, 170 ont accepté et 120 ont finalement recruté.

#### **Echecs de recrutement**

L'essai Event dans la thrombose veineuse profonde a enrôlé 7 patients sur les 250 prévus, l'essai Rhinosinusitis a enrôlé 45 patients sur les 400 prévus. Deux essais en cours présentent actuellement des difficultés de recrutement.

### Facteurs associés au recrutement

Les facteurs fréquemment associés à des difficultés de recrutement sont le recrutement via des médecins spécialistes, le manque d'implication des médecins thermaux et le manque de connaissance sur le recours à la cure thermale pour la pathologie.

Les facteurs fréquemment associés à un meilleur recrutement sont le recrutement direct via les médias, la fréquence de la pathologie étudiée, la bonne connaissance de la cure thermale pour la pathologie, l'implication des médecins thermaux, un accès aisé à l'investigateur pour l'inclusion.

### *Résultats des ECRs conduits sous l'égide de l'AFRETH*

Tous les ECRs ont montré une différence statistiquement significative pour le critère de jugement principal sauf l'essai Thermes&veines. L'importance de l'effet traitement variait selon les pathologies.

### **Conclusion**

La médecine thermale a démontré qu'elle constituait une option thérapeutique légitime en raison des résultats observés dans les essais randomisés. L'objectif est de continuer d'évaluer les bénéfices des cures thermales notamment pour les pathologies où le produit thermominéral est directement au contact des lésions (affections dermatologiques et respiratoires). La principale difficulté est de recruter un nombre suffisant de malades pour évaluer le bénéfice des cures thermales ce qui souligne l'importance de développer de nouveaux paradigmes d'évaluation. Une possibilité pourrait être d'utiliser les données de l'assurance maladie et notamment d'avoir accès à l'Echantillon Généraliste de Bénéficiaires (EGB) pour savoir quelles informations sont disponibles concernant les cures thermales en particulier l'indication et le lieu de la cure. L'EGB est un échantillon permanent représentatif de la population couverte par l'Assurance maladie française. Il contient des informations anonymes sur les caractéristiques sociodémographiques et médicales des bénéficiaires et les prestations qu'ils ont perçues. L'EGB résulte d'un sondage au 1/97e sur le numéro de sécurité sociale (NIR) des bénéficiaires de l'Assurance maladie française, qu'ils aient ou non perçu des remboursements de soins. Il regroupe actuellement près de 500 000 bénéficiaires du régime des travailleurs salariés autres que les fonctionnaires et les étudiants (régime général). Si les informations sur l'EGB sont suffisantes, on pourrait évaluer des critères de jugement portant sur la consommation de soins (médicaments et séances de rééducation) ainsi que les hospitalisations.

## Points importants et particularités méthodologiques dans l'évaluation des interventions complexes

### ***Définitions et typologie***

#### Traitement non-pharmacologique

Un traitement non-pharmacologique est défini comme un traitement ou une intervention n'impliquant pas principalement ni uniquement l'utilisation de substances pharmacologiques.

#### Intervention complexe

Une intervention complexe est une intervention constituée d'un certain nombre de composants, qui peuvent agir à la fois de façon indépendante et interdépendante.

Les interventions non-pharmacologiques sont très nombreuses et très variées.

L'évaluation thérapeutique des traitements non-pharmacologiques présente des caractéristiques communes avec celle des traitements pharmacologiques et des caractéristiques spécifiques(24).

Selon le Cochrane handbook(25), les différents biais pouvant affecter les résultats d'un essai randomisé contrôlé sont :

- Biais de sélection qui désignent des différences systématiques entre les caractéristiques initiales des groupes à comparer. Pour limiter le risque de biais de sélection, la méthode pour générer la séquence de randomisation doit être adéquate (séquence informatique, table de nombres aléatoires, tirage à pile ou face) et le principe d'assignation secrète respectée (le résultat de la randomisation doit être imprévisible lors de l'inclusion).
- Biais de performance qui désignent des différences systématiques entre les groupes dans la prise en charge des participants. Pour limiter le risque de biais de performance, les participants et l'équipe soignante doivent être en aveugle des interventions reçues.
- Biais de détection qui désignent des différences systématiques entre les groupes dans la manière d'évaluer le critère de jugement. Pour limiter le risque de biais de détection, l'évaluateur doit être en aveugle si le critère de jugement est subjectif.
- Biais d'attrition qui désignent des différences systématiques entre les groupes concernant des retraits des participants de l'essai. Pour limiter le risque de biais d'attrition, l'analyse doit être en intention de traiter et une stratégie de gestion des données manquantes doit être développée.
- Biais de publication et de présentation sélective des critères de jugement qui désignent des différences systématiques entre les résultats rapportés et les résultats non-rapportés et qui sont liés à la publication sélective des essais rapportant des résultats statistiquement significatifs ou à la présentation

sélective des critères de jugement avec des résultats statistiquement significatifs

Les caractéristiques communes à l'évaluation des interventions pharmacologiques et non-pharmacologiques sont une méthode adéquate pour générer la séquence de randomisation, une assignation secrète, une analyse en intention de traiter ainsi que la publication des résultats de l'étude telle qu'elle a été planifiée.

## ***Particularités méthodologiques de l'évaluation des traitements non-pharmacologiques***

### ***Aveugle***

L'aveugle est beaucoup plus compliqué à obtenir pour un traitement non-pharmacologique que pour un traitement pharmacologique. Il est important pour limiter les biais surtout si le critère de jugement est subjectif. Une étude méta-épidémiologique a ainsi montré que l'effet traitement était surestimé de 25% en moyenne dans les essais non en aveugle par rapport aux essais en aveugle(26).

### Faisabilité de l'aveugle

Le premier point à évaluer est la faisabilité de l'aveugle. Une étude méthodologique a montré que l'aveugle des patients était faisable dans 42% des essais portant sur des traitements non-pharmacologiques, l'aveugle des thérapeutes n'était faisable que dans 12% des essais et l'aveugle de l'évaluateur n'était faisable que dans 34% des essais(27). Pour l'évaluateur, il faut prendre en compte le type de critère de jugement. S'il s'agit d'un « patient reported outcome » c'est-à-dire un critère évalué par le patient, l'évaluation ne peut être en aveugle que si le patient est en aveugle.

### Comment obtenir l'aveugle ?

L'aveugle est relativement simple à obtenir quand il s'agit de traitements pharmacologiques mais beaucoup plus compliqué à obtenir quand il s'agit de traitements non-pharmacologiques(28, 29).

Il existe 3 possibilités pour obtenir l'aveugle

- 1) Aveugle complet : obtenu avec un placebo ou par simulation de l'intervention et absence de contact avec le thérapeute ayant réalisé l'intervention. La question d'un placebo de la cure thermale n'est pas simple : il pourrait s'agir d'un séjour dans une station thermale sans cure ou d'une intervention de type « attention control intervention » avec des sessions de même fréquence et durée mais un contenu différent. Cela sous estime l'effet traitement.
- 2) Aveugle partiel : On manipule l'information donnée aux patients. Soit les patients ne sont pas informés de l'existence d'un placebo soit ils ne sont pas informés des hypothèses de l'essai (insu des hypothèses).
- 3) Evaluation en aveugle du critère de jugement : par exemple, évaluation centralisée de photos dans le psoriasis. Cela n'est possible que si le critère de jugement n'est pas un « Patient reported outcome ».

### Choix du comparateur

Le choix du comparateur dépend de ce qu'on veut mesurer.

L'effet traitement résulte de la combinaison de plusieurs éléments : l'effet Hawthorne, l'effet placebo, les effets non liés au traitement (régression vers la moyenne, évolution naturelle de la maladie, caractéristiques des patients, erreur de mesure).

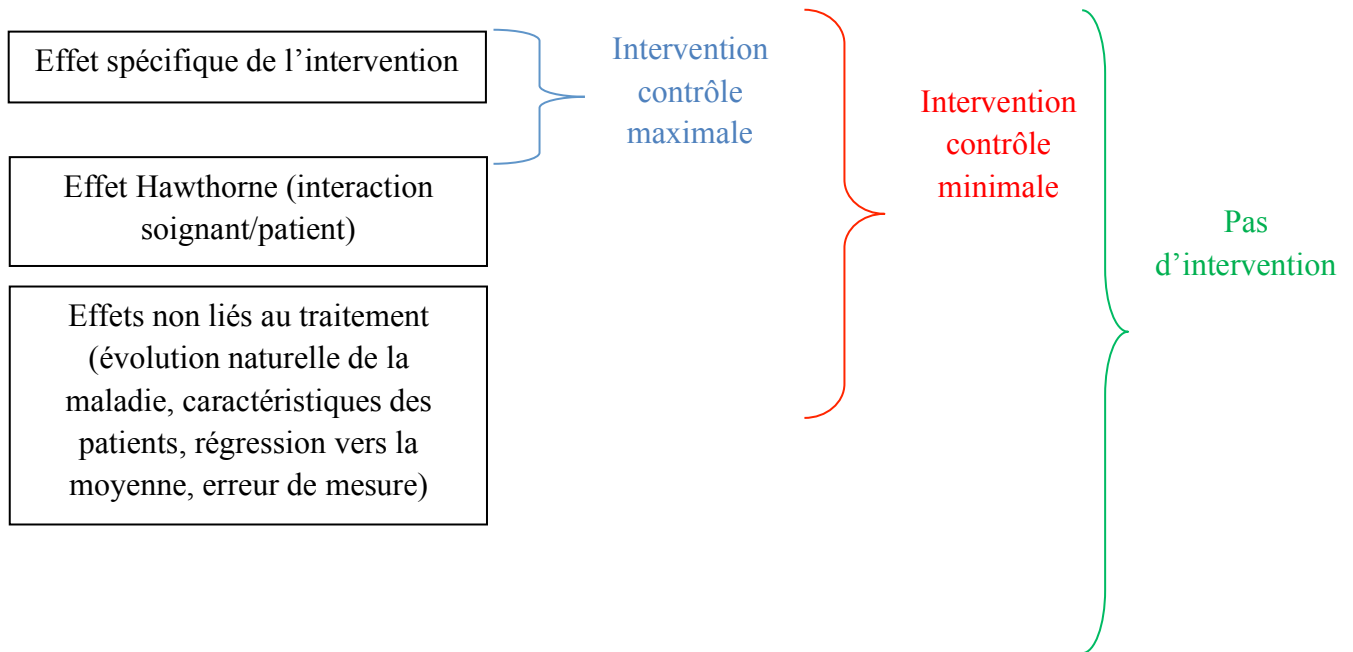


Figure 6: Quel effet traitement veut-on mesurer ?(30)

### Particularités liées au caractère complexe de l'intervention

La cure thermique est une intervention non-pharmacologique complexe composée de multiples composantes

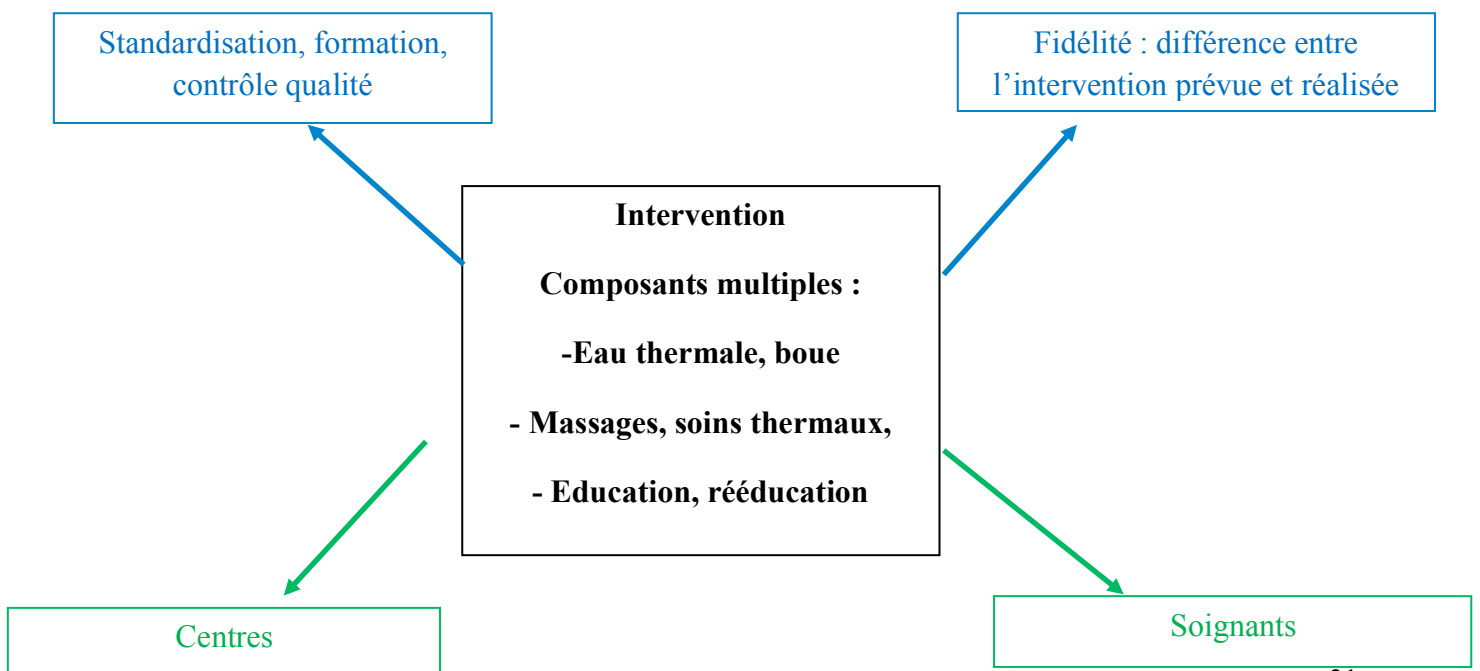


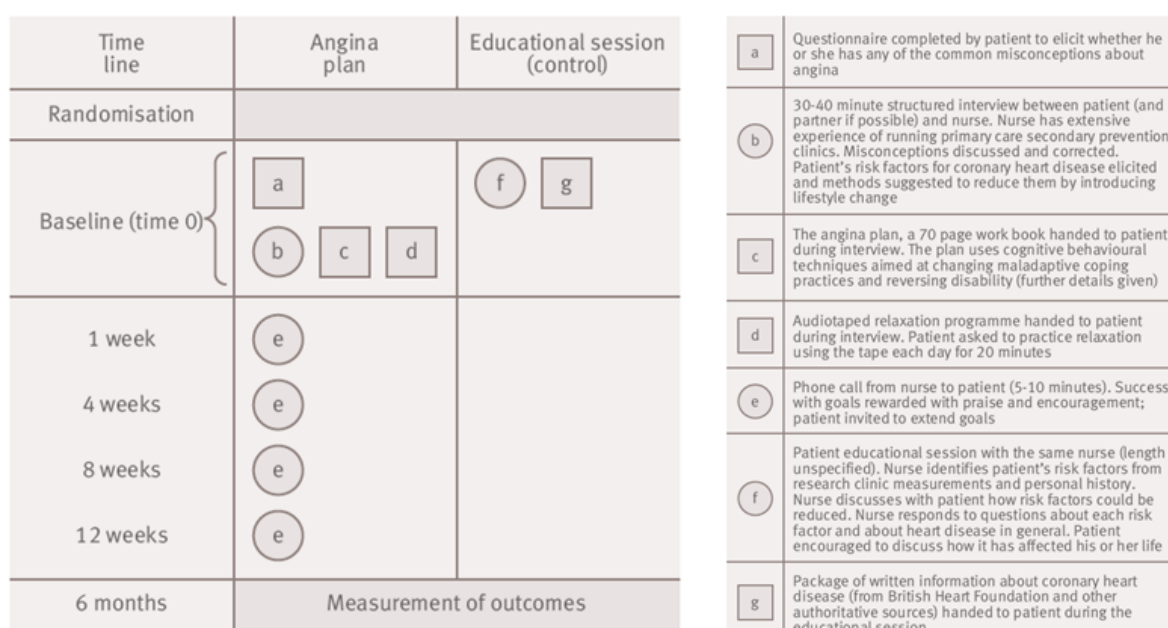
Figure 7 :Particularités méthodologiques d'une intervention complexe

### Standardisation

La manière dont on décrit une intervention non-pharmacologique est totalement insuffisante. D'après Glasziou(31), seulement 30% des interventions non-pharmacologiques peuvent être reproduites d'après les données de l'article. L'intervention peut également varier d'un centre à l'autre. Il faut standardiser suffisamment l'intervention pour qu'elle puisse être reproduite dans d'autres centres. L'individualisation du traitement est beaucoup plus fréquente pour les traitements non-pharmacologiques que pharmacologiques. La standardisation n'est pas contradictoire avec l'individualisation du traitement. Il faudra standardiser et décrire la personnalisation.

L'utilisation de pictogrammes peut être utile pour standardiser et décrire l'intervention(32).

Un exemple est donné ci-après.



Perera, R. et al. BMJ 2007;334:127-129

Figure 8: Exemple de graphique permettant de décrire une intervention standardisée

### Compliance vis à vis de l'intervention

Evaluer la compliance consiste à évaluer s'il y a eu des différences entre l'intervention prévue et celle réalisée. La compliance peut s'évaluer à plusieurs niveaux : au niveau du patient, du thérapeute et du centre.

### Influence du thérapeute et des centres

### Effet volume

Le thérapeute et le centre ont une influence beaucoup plus importante pour les interventions complexes. Il y a en effet une relation importante entre le volume d'activité d'un thérapeute ou d'un centre et les résultats en raison de la complexité de l'intervention. De nombreuses études ont montré cet effet volume(33, 34) et il semble que l'élément le plus important soit le volume du centre. La spécialisation semble également avoir un rôle important(35).

### Courbe d'apprentissage

Un paramètre à prendre également en compte quand on évalue une nouvelle intervention complexe est la courbe d'apprentissage du thérapeute. L'apprentissage d'une nouvelle technique peut s'accompagner au début d'erreurs et d'effets secondaires ce qui peut biaiser le résultat en défaveur de la nouvelle intervention(36).

### « Clustering effect »

Quand on évalue une intervention complexe, les données concernant les résultats seront plus corrélées pour les patients au sein d'un même centre que pour des patients entre 2 centres différents. Il faudrait théoriquement le prendre en compte lors de la planification de l'essai et du calcul d'effectif car le « clustering effect » peut entraîner un manque de puissance et le prendre en compte dans l'analyse. Pour cela, il faut avoir une idée du coefficient de corrélation intraclasse ce qui n'est pas toujours possible.

# Recommandations pour évaluer une intervention complexe

## 1) Concernant l'intervention

- Identifier les différentes composantes de l'intervention prévue
- Utiliser des méthodes de standardisation
- Evaluer l'adhérence à cette intervention

## 2) Concernant les thérapeutes

- Définir les critères d'éligibilité des thérapeutes
- Définir le nombre de thérapeutes
- Définir le nombre d'interventions réalisées par chaque thérapeute
- Définir leur expérience
- Former les thérapeutes et évaluer leur courbe d'apprentissage

## 3) Concernant les centres

- Définir les critères d'éligibilité des centres
- Définir le type de centre
- Identifier les infrastructures nécessaires
- Définir l'expérience du centre (volume)
- Définir le nombre de malades vus dans chaque centre

## 4) Prendre en compte le « clustering effect »

- Lors de la planification de l'essai
- Lors de l'analyse statistique



## **Conséquences pratiques pour l'évaluation thérapeutique des cures thermales**

Les cures thermales concernent 500 000 personnes par an en France. Les principales indications sont l'arthrose et l'insuffisance veineuse chronique. La cure thermale apparaît comme une thérapeutique à visée symptomatique permettant d'améliorer les symptômes et la qualité de vie de patients souffrant de pathologies chroniques douloureuses retentissant sur leurs activités quotidiennes et pour lesquels la prise en charge habituelle est en échec. Il s'agit d'abord d'un service médical rendu à la personne. Les cures thermales répondent à un besoin de santé publique non couvert par la prise en charge habituelle ce qu'on appelle le fardeau résiduel.

### ***Réflexion sur les critères de jugement***

#### ***Quels types de critère de jugement***

L'évaluation du bénéfice des cures thermales devrait porter sur des critères de jugement fonctionnels visant à évaluer le confort des patients. Un critère de jugement composite permettrait d'appréhender différentes dimensions importantes pour le patient. De manière générale, beaucoup d'évaluation portent sur la mise en place, la faisabilité et pas assez sur des critères plus pertinents pour le patient.

La stratégie serait donc d'évaluer des critères de jugement pertinents pour le patient tels que son niveau de douleur, son confort sa qualité de vie qui sont des « patient reported outcomes » et d'utiliser des critères de jugement de type médico-économique en critères secondaires comme la diminution de consommation d'AINS ou de kinésithérapie.

Ces critères doivent être également simples, disponibles pour la majorité des patients et doivent parler aux patients, médecins et décideurs.

Les critères de jugement pertinents peuvent différer en fonction de l'indication de la cure thermale. Il faudrait établir par pathologie les critères de jugement pertinents à évaluer (consensus d'experts) et évaluer l'importance de l'effet traitement pour ces critères dans chacune des indications de la cure thermale. Par exemple, dans l'arthrose du genou, les critères pertinents sont la douleur, les incapacités, ou le fait de retarder la mise en place d'une prothèse.

#### ***Horizon temporel***

Quand on définit un critère de jugement, on doit également définir son horizon temporel. Il convient donc d'avoir une réflexion sur l'horizon temporel qu'on souhaite évaluer: est ce qu'on veut évaluer les bénéfices de la cure thermale en sortie de cure, à 6 mois, à 1 an ou à plus long terme ? Il est intéressant de savoir si les

bénéfices observés sont durables dans le temps. Les études de cohorte peuvent être intéressantes pour évaluer les bénéfices à long terme.

### ***Evaluation des évènements indésirables***

La cure thermale est habituellement considérée comme une thérapeutique sans risque. Néanmoins, il ne faut pas négliger l'évaluation de possibles évènements indésirables lors de la cure thermale. Il y a eu des cas de légionellose.

## ***Réflexion sur ce qu'on souhaite évaluer***

### ***Quelle intervention ?***

Faut-il évaluer la cure thermale globalement ? Ou évaluer quelles composantes sont efficaces notamment l'eau thermale de par sa composition. La cure thermale est une intervention complexe dont l'eau thermale ne représente qu'une petite partie. La cure thermale est issue d'une longue tradition et fait intervenir des massages par des masseurs qualifiés. Le fait d'être également dans une station thermale avec son environnement particulier est également important. Il faut différencier à ce titre les gens qui viennent de loin pour faire une cure (75% des curistes) de ceux habitant dans la région (25%). Dans le 1<sup>er</sup> cas, on évalue le séjour thermal avec repos et le fait de se sentir en vacances. Dans le second cas, les patients font la cure en plus de leurs activités quotidiennes.

Ce qui semble important c'est l'évaluation de la cure thermale de manière globale (l'eau, les massages, les interlocuteurs) c'est-à-dire l'effectiveness.

Est-ce que si ça marche de manière globale, ça vaut le coup d'évaluer ce qui marche par exemple l'eau thermale qui, en raison de sa minéralité, a des propriétés physiques et chimiques. Des études en double aveugle ont déjà portées sur l'évaluation de l'eau thermale et ont montré un effet analgésique ainsi que des effets physiologiques (augmentation du taux d'hormones et de la circulation périovarienne). Quand on évalue une intervention chirurgicale c'est l'intervention de manière globale qui est évaluée et on ne se demande pas ce qui marche dedans. De plus, si l'évaluation ne montre pas de bénéfice de manière globale, ça ne sert à rien d'aller plus loin.

### ***Quelle durée ?***

La durée d'une cure thermale en France est actuellement de 3 semaines et l'assurance maladie rembourse les cures thermales d'une durée de 3 semaines. Il y a des pays où la cure thermale ne dure que 2 semaines. Il serait intéressant d'évaluer 2 semaines voire une semaine de cure par rapport à 3 semaines. Mais la première étape est d'évaluer les bénéfices de 3 semaines de cure.

### ***Standardisation de l'intervention***

Il semble que le contenu des séances soit bien standardisé. Les différentes composantes de l'intervention (description de la durée de séance, des activités réalisées au cours des séances, de la formation des thérapeutes,...) doivent être bien décrites.

## ***Recrutement des patients***

### ***Difficultés de recrutement***

Les difficultés de recrutement ne sont pas spécifiques des cures thermales mais majorées par le fait que la cure thermale nécessite un aménagement de son emploi du temps que tout le monde n'est pas prêt à faire en raison de sa durée.

Le recrutement à partir d'un volontariat des personnes via des annonces dans la presse marche généralement bien. Il semble plus difficile quand il fait appel aux médecins spécialistes ou généralistes qui n'ont pas beaucoup de temps à consacrer au recrutement des patients lors de leur consultation et qui n'étant pas directement impliqués dans la cure thermale sont généralement moins motivés. Ce qui est faisable, c'est de mettre des affiches pour informer les patients dans la salle d'attente des médecins ou dans les pharmacies.

Les modalités de recrutement peuvent différer d'une pathologie à l'autre. Il n'y a pas de paradigme : certaines modalités de recrutement vont être adaptées dans certaines pathologies mais moins dans d'autres. Une autre possibilité peut être de passer par les associations de malades.

Le recrutement dépend également de l'investissement de l'investigateur principal et des personnes qui assurent le monitoring.

### ***Problème de représentativité des patients***

Certains modes de recrutement risquent de sélectionner des personnes particulières qui ne sont pas forcément représentatives de la population habituelle de curistes par exemple le recrutement via des annonces dans la presse. La question de la représentativité des patients se pose toujours quelque soit le type d'étude et particulièrement pour les ECRs. De nombreuses études méthodologiques montrent que la population incluse dans les essais randomisés contrôlés n'est absolument pas représentative de la population traitée dans la vraie vie. Il faut également noter que les essais ne portent pas forcément sur la population de curistes habituels : certains portent principalement sur des patients venant des alentours qui représentent habituellement 25% des curistes.

### ***Sur quelles bases sélectionne-t-on des patients atteints d'une certaine pathologie?***

La tendance générale actuelle est de se baser sur le diagnostic déclaré par les patients quelque soit la pathologie(37, 38). Cela permet de diminuer le coût et s'intègre dans le concept du « large simple pragmatic trial »(39).

### *Utilisation des nouveaux outils technologiques*

L'essor d'internet et le développement de nouveaux outils technologiques pourraient faciliter le recrutement des patients. On pourrait ainsi imaginer un recrutement des patients via Facebook ou Twitter. Certains outils permettent de repérer les gens qui font des recherches sur internet concernant certaines pathologies (service payant de Google). En prenant exemple sur l'étude Nutrinet santé ou l'Army of Women aux Etats-Unis, on pourrait imaginer des cohortes de patients dans certaines pathologies avec un site internet permettant aux gens de rentrer des données sur eux-mêmes concernant leur pathologie et que ces personnes soient ainsi contactées par email pour repérer les gens potentiellement intéressés pour faire une cure thermale. Le Pr Serge Hercberg et son équipe a montré une bonne représentativité des patients pour la cohorte d'internautes Nutrinet(40). Cela pourrait marcher pour la fibromyalgie car les patients sont relativement jeunes mais peut être difficile à mettre en place dans des pathologies où les patients sont âgés.

### *Utilisation des bases de données médico-administratives*

D'autres possibilités sont d'utiliser les bases de données de l'assurance maladie ou éventuellement les logiciels de gestion des patients des cabinets de médecine générale.

Concernant les données de l'assurance maladie, une première étape serait d'étudier l'EGB. Si l'étude préliminaire sur l'EGB montre que l'on dispose de suffisamment d'informations concernant les cures thermales, une demande d'accès au SNIIRAM pourrait être envisagée par la suite. Ces bases de données sont d'un intérêt majeur car elles sont représentatives de la population française. On pourra croiser les informations pour avoir la consommation de soins des curistes.

## Propositions d'amélioration dans le cadre de la mise en place de projets visant à évaluer les cures thermales

Points méthodologiques	Recommandations	Etape préliminaire
Critères de jugement principaux	Pertinents pour le patient Aspect fonctionnel : confort, qualité de vie	Etablissement des critères pertinents à évaluer pathologie par pathologie
Intervention évaluée	- La cure thermique dans son ensemble c'est-à-dire l' « effectiveness » - Standardisation et description des différentes composantes	Identification des différentes composantes de l'intervention
Pallier aux difficultés de recrutement Problèmes de représentativité des patients inclus dans les essais Suivi à long terme Evènements indésirables	Utiliser les données de l'assurance maladie représentatives de la population française	Consultation de l'EGB pour savoir si informations détaillées sur les cures thermales

# Références

1. Zelen M. A new design for randomized clinical trials. *N Engl J Med*. 1979 May 31;300(22):1242-5.
2. Adamson J, Cockayne S, Puffer S, Torgerson DJ. Review of randomised trials using the post-randomised consent (Zelen's) design. *Contemp Clin Trials*. 2006 Aug;27(4):305-19.
3. Schellings R, Kessels AG, ter Riet G, Knottnerus JA, Sturmans F. Randomized consent designs in randomized controlled trials: systematic literature search. *Contemp Clin Trials*. 2006 Aug;27(4):320-32.
4. Brewin CR, Bradley C. Patient preferences and randomised clinical trials. *BMJ*. 1989 Jul 29;299(6694):313-5.
5. Relton C, Torgerson D, O'Cathain A, Nicholl J. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. *BMJ*. 2010;340:c1066.
6. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis*. 1967 Aug;20(8):637-48.
7. Hewitt CE, Torgerson DJ, Miles JN. Is there another way to take account of noncompliance in randomized controlled trials? *CMAJ*. 2006 Aug 15;175(4):347.
8. Sackett DL. *Evidence-Based Medicine: How to practice and to teach EBM*. NY: Churchill Livingstone. 1997.
9. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000 Jun 22;342(25):1878-86.
10. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000 Jun 22;342(25):1887-92.
11. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Med*. 2011 May;8(5):e1001026.
12. 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 Aug 15;286(7):821-30.
13. Grodstein F, Clarkson TB, Manson JE. Understanding the divergent data on postmenopausal hormone therapy. *N Engl J Med*. 2003 Feb 13;348(7):645-50.
14. von Elm E, Egger M. The scandal of poor epidemiological research. *BMJ*. 2004 Oct 16;329(7471):868-9.
15. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess*. 1998;2(13):i-iv, 1-124.
16. Mant D. Can randomised trials inform clinical decisions about individual patients? *Lancet*. 1999 Feb 27;353(9154):743-6.
17. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effect. *Biometrika*. 1983;70:41-55.
18. Weintraub WS, Grau-Sepulveda MV, Weiss JM, O'Brien SM, Peterson ED, Kolm P, et al. Comparative effectiveness of revascularization strategies. *N Engl J Med*. 2012 Apr 19;366(16):1467-76.
19. 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 Feb 1;163(3):262-70.
20. Feinstein A. The role of observational studies in the evaluation of therapy. *Stat Med*. 1984 Oct-Dec;3(4):341-5.

21. Dubois O, Salamon R, Germain C, Poirier MF, Vaugeois C, Banwarth B, et al. Balneotherapy versus paroxetine in the treatment of generalized anxiety disorder. *Complement Ther Med*. 2010 Feb;18(1):1-7.
22. Forestier R, Desfour H, Tessier JM, Francon A, Foote AM, Genty C, et al. Spa therapy in the treatment of knee osteoarthritis: a large randomised multicentre trial. *Ann Rheum Dis*. 2010 Apr;69(4):660-5.
23. Hanh T, Serog P, Fauconnier J, Batailler P, Mercier F, Roques CF, et al. One-year effectiveness of a 3-week balneotherapy program for the treatment of overweight or obesity. *Evid Based Complement Alternat Med*. 2012;2012:150839.
24. 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 Aug 27;290(8):1062-70.
25. Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions version 5.0.0*. Cochrane Collaboration. 2008.
26. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*. 2008 Mar 15;336(7644):601-5.
27. 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 Jun;57(6):543-50.
28. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hrobjartsson A, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med*. 2006 Oct;3(10):e425.
29. Boutron I, Guittet L, Estellat C, Moher D, Hrobjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med*. 2007 Feb;4(2):e61.
30. Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ*. 2005 May 21;330(7501):1202-5.
31. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ*. 2008 Jun 28;336(7659):1472-4.
32. Perera R, Heneghan C, Yudkin P. Graphical method for depicting randomised trials of complex interventions. *BMJ*. 2007 Jan 20;334(7585):127-9.
33. Cebul RD, Snow RJ, Pine R, Hertzner NR, Norris DG. Indications, outcomes, and provider volumes for carotid endarterectomy. *JAMA*. 1998 Apr 22-29;279(16):1282-7.
34. 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 Sep 17;137(6):511-20.
35. Hagen TP, Vaughan-Sarrazin MS, Cram P. Relation between hospital orthopaedic specialisation and outcomes in patients aged 65 and older: retrospective analysis of US Medicare data. *BMJ*. 2010;340:c165.
36. Tekkis PP, Senagore AJ, Delaney CP, Fazio VW. Evaluation of the learning curve in laparoscopic colorectal surgery: comparison of right-sided and left-sided resections. *Ann Surg*. 2005 Jul;242(1):83-91.
37. Cascade E, Marr P, Winslow M, Burgess A, Nixon M. Conducting research on the Internet: medical record data integration with patient-reported outcomes. *J Med Internet Res*. 2012;14(5):e137.
38. Rasooly I, Papageorgiou AC, Badley EM. Comparison of clinical and self reported diagnosis for rheumatology outpatients. *Ann Rheum Dis*. 1995 Oct;54(10):850-2.
39. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med*. 1984 Oct-Dec;3(4):409-22.
40. Touvier M, Mejean C, Kesse-Guyot E, Pollet C, Malon A, Castetbon K, et al. Comparison between web-based and paper versions of a self-administered anthropometric questionnaire. *Eur J Epidemiol*. 2010 May;25(5):287-96.

