

# AVIS

Notre référence: 03-102451-780

## L'adoption pour l'ICH<sup>1</sup> ligne directrice

Santé Canada a le plaisir d'annoncer l'adoption de cette ligne directrice de l'ICH:

### **E9: Principes statistiques pour les essais cliniques**

Cette ligne directrice a été élaborée par un groupe d'experts de l'ICH et a fait l'objet de consultations, menées par les organismes de réglementation, conformément au processus de l'ICH. Le Comité directeur de l'ICH en a approuvé la version finale et en a recommandé l'adoption par les organismes de réglementation de l'Union européenne, du Japon et des États-Unis.

En adoptant cette ligne directrice de l'ICH, Santé Canada fait siens les principes et les pratiques qui y sont énoncés. Cette document doit être lu en parallèle avec cette d'avis d'accompagnement et les sections pertinentes des autres lignes directrices applicables du Santé Canada.

Plusieurs lignes directrices, incluant celle-ci, sont disponibles sur le site Internet de la **Direction des produits thérapeutiques / Direction des produits biologiques et thérapies génétiques** (<http://www.hc-sc.gc.ca/hpb-dgps/therapeut>). Pour accéder à la liste des "copies papier" des lignes directrices disponibles, veuillez consulter la liste qui apparaît sur les bons de commande des publications et des directives (publiés sur le site Internet de la DPT/DPBTG), ou veuillez communiquer avec le coordonnateur / coordonnatrice des publications<sup>2</sup>.

Si vous avez des questions concernant cette ligne directrice, veuillez communiquer avec:

Robert K.Y. Li  
Gestionnaire intérimaire  
Gestion des risques et méthodes  
Bureau de la politique  
Direction des produits thérapeutiques  
Santé Canada  
Holland Cross, Tour "B"  
2ième étage, I.A. 3102C3  
1600 rue Scott  
Ottawa (Ontario)  
K1A 1B6

courrier électronique: [policy\\_bureau\\_enquiries@hc-sc.gc.ca](mailto:policy_bureau_enquiries@hc-sc.gc.ca)

Téléphone: (613) 941-3693

Télécopieur : (613) 941-5035

---

<sup>1</sup> *International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use*

<sup>2</sup> Tel: (613) 954-6466; courrier électronique: [publications\\_coordinator@hc-sc.gc.ca](mailto:publications_coordinator@hc-sc.gc.ca)



# LIGNE DIRECTRICE À L'INTENTION DE L'INDUSTRIE

Principes statistiques pour les essais cliniques  
ICH thème E9

Publication autorisée par le  
ministre de la Santé

Date d'approbation	2003/02/10
Date mis en vigueur	2003/02/10

**Direction générale des produits de santé et des aliments**

<p>Notre Mission est d'aider les Canadiens et les Canadiennes à maintenir et à améliorer leur état de santé.</p> <p style="text-align: right;"><i>Santé Canada</i></p>	<p>Le mandat de la DGPSA est d'adopter une approche intégrée à la gestion des risques et des avantages pour la santé liés aux produits de santé et aux aliments:</p> <ul style="list-style-type: none"> <li>• en réduisant les facteurs de risque pour la santé des Canadiens et des Canadiennes tout en maximisant la protection offerte par le système réglementaire des produits de la santé et des aliments;</li> <li>• et en favorisant des conditions qui permettent aux Canadiens et aux Canadiennes de faire des choix sains ainsi qu'en leur donnant des renseignements afin qu'ils ou qu'elles puissent prendre des décisions éclairées en ce qui a trait à leur santé.</li> </ul> <p style="text-align: right;"><i>Direction générale des produits de santé et des aliments</i></p>
--	--

**LE SITE WEB DE LA DIRECTION DES PRODUITS THÉRAPEUTIQUES / DIRECTION DES PRODUITS BIOLOGIQUES ET DES THÉRAPIES GÉNÉTIQUES**

**LAISSEZ VOTRE ORDINATEUR FAIRE LES RECHERCHES!**

...Vous voulez savoir comment commercialiser un nouveau médicament?

...Vous souhaitez obtenir des renseignements au sujet du processus de réglementation des médicaments?

...Vous voulez connaître quels sont les médicaments les plus récemment autorisés au Canada?

...Vous souhaitez avoir un accès direct à nos formulaires et à nos politiques?

...Vous voulez connaître quelles sont les contraintes en matière d' étiquetage des médicaments?

Vous pouvez obtenir ces renseignements et plusieurs autres en consultant

**le site web de la direction des produits thérapeutiques / direction des produits biologiques et des thérapies génétiques**

à

[www.hc-sc.gc.ca/hpb-dgps/therapeut](http://www.hc-sc.gc.ca/hpb-dgps/therapeut)

© Ministre, Travaux publics et services gouvernementaux Canada 2003

Disponible au Canada par l'entremise de  
Santé Canada - Publications  
Edifice Brooke Claxton, L. A. #0913A  
Pré Tunney  
OTTAWA (Ontario)  
K1A 0K9

téléphone : (613) 954-5995  
télécopieur : (613) 941-5366

*also available in English under the following Title:* Statistical Principles for Clinical Trials

N° de catalogue H49-171/2003F  
ISBN 0-662-88244-X

## AVANT-PROPOS

La présente ligne directrice a été élaborée par un groupe d'experts de l'ICH et a fait l'objet de consultations, menées par les organismes de réglementation, conformément au processus de l'ICH. Le Comité directeur de l'ICH en a approuvé la version finale et en a recommandé l'adoption par les organismes de réglementation de l'Union européenne, du Japon et des États-Unis.

En adoptant cette ligne directrice de l'ICH, Santé Canada fait siens les principes et les pratiques qui y sont énoncés. Ce document devrait être lu en parallèle avec l'avis d'accompagnement et les sections pertinentes des autres lignes directrices qui s'appliquent.

Les lignes directrices sont des documents destinés à guider l'industrie et les professionnels de la santé sur la **façon** de se conformer aux politiques et aux lois et règlements qui régissent leurs activités. Elles servent également de guide au personnel lors de l'évaluation et de la vérification de la conformité et permettent ainsi d'appliquer les mandats d'une façon équitable, uniforme et efficace.

Les lignes directrices sont des outils administratifs n'ayant pas force de loi, ce qui permet une certaine souplesse d'approche. Les principes et les pratiques énoncés dans le présent document **pourraient être** remplacés par d'autres approches, à condition que celles-ci s'appuient sur une justification scientifique adéquate. Ces autres approches devraient être examinées préalablement en consultation avec le programme concerné pour s'assurer qu'elles respectent les exigences des lois et des règlements applicables.

Corollairement à ce qui précède, il importe également de mentionner que Santé Canada se réserve le droit de demander des renseignements ou du matériel supplémentaire, ou de définir des conditions dont il n'est pas explicitement question dans la ligne directrice, et ce, afin que le ministère puisse être en mesure d'évaluer adéquatement l'innocuité, l'efficacité ou la qualité d'un produit thérapeutique donné. Santé Canada s'engage à justifier de telles demandes et à documenter clairement ses décisions.

## TABLE DES MATIÈRES

1.	INTRODUCTION .....	<u>1</u>
1.1	Contexte et objet .....	<u>1</u>
1.2	Champ d'application et orientation .....	<u>2</u>
2.	CONSIDÉRATIONS RELATIVES À LA MISE AU POINT CLINIQUE DES MÉDICAMENTS .....	<u>4</u>
2.1	Contexte des essais .....	<u>4</u>
2.1.1	<i>Plan de mise au point</i> .....	<u>4</u>
2.1.2	<i>Essai confirmatoire</i> .....	<u>5</u>
2.1.3	<i>Essai exploratoire</i> .....	<u>6</u>
2.2	Champ d'application des essais .....	<u>6</u>
2.2.1	<i>Population</i> .....	<u>6</u>
2.2.2	<i>Variables primaires et secondaires</i> .....	<u>7</u>
2.2.3	<i>Variables composées</i> .....	<u>8</u>
2.2.4	<i>Variables d'évaluation globale</i> .....	<u>9</u>
2.2.5	<i>Variables primaires multiples</i> .....	<u>10</u>
2.2.6	<i>Variables de substitution</i> .....	<u>10</u>
2.2.7	<i>Variables qualitatives catégorielles</i> .....	<u>11</u>
2.3	Méthodes de réduction du biais .....	<u>12</u>
2.3.1	<i>Insu</i> .....	<u>12</u>
2.3.2	<i>Randomisation</i> .....	<u>14</u>
3.	CONSIDÉRATIONS RELATIVES AU PLAN DE L'ESSAI .....	<u>17</u>
3.1	Configuration du plan .....	<u>17</u>
3.1.1	<i>Plan à groupes parallèles</i> .....	<u>17</u>
3.1.2	<i>Plan croisé</i> .....	<u>17</u>
3.1.3	<i>Plans factoriels</i> .....	<u>18</u>
3.2	Essais multicentriques .....	<u>19</u>
3.3	Type de comparaison .....	<u>22</u>
3.3.1	<i>Essais visant à montrer une relation de supériorité</i> .....	<u>22</u>
3.3.2	<i>Essais visant à montrer une relation d'équivalence ou de non-infériorité</i> .....	<u>22</u>
3.3.3	<i>Essais visant à montrer une relation dose-effet</i> .....	<u>25</u>
3.4	Essais séquentiels par groupes .....	<u>25</u>
3.5	Taille de l'échantillon .....	<u>26</u>
3.6	Saisie et traitement des données .....	<u>28</u>

4.	CONSIDÉRATIONS RELATIVES À LA CONDUITE DES ESSAIS .....	<a href="#">28</a>
4.1	Surveillance des essais et analyse intermédiaire .....	<a href="#">28</a>
4.2	Modification des critères d'inclusion et d'exclusion .....	<a href="#">29</a>
4.3	Taux de constitution de l'échantillon .....	<a href="#">29</a>
4.4	Correction de la taille de l'échantillon .....	<a href="#">30</a>
4.5	Analyse intermédiaire et interruption du traitement .....	<a href="#">30</a>
4.6	Rôle du Comité indépendant de surveillance des données (CISD) .....	<a href="#">32</a>
5.	CONSIDÉRATIONS RELATIVES À L'ANALYSE DES DONNÉES .....	<a href="#">33</a>
5.1	Définition préalable de l'analyse .....	<a href="#">33</a>
5.2	Ensembles d'analyse .....	<a href="#">34</a>
5.2.1	<i>Ensemble d'analyse intégral</i> .....	<a href="#">34</a>
5.2.2	<i>Ensemble conforme au protocole</i> .....	<a href="#">37</a>
5.2.3	<i>Fonctions des divers ensembles d'analyse</i> .....	<a href="#">38</a>
5.3	Valeurs manquantes et valeurs aberrantes .....	<a href="#">38</a>
5.4	Transformation de données .....	<a href="#">39</a>
5.5	Estimation, intervalles de confiance et tests d'hypothèses .....	<a href="#">40</a>
5.6	Modification du niveau de signification et du niveau de confiance .....	<a href="#">41</a>
5.7	Sous-groupes, interactions et covariables .....	<a href="#">41</a>
5.8	Intégrité des données et validité des logiciels .....	<a href="#">42</a>
6.	ÉVALUATION DE L'INNOCUITÉ ET DE LA TOLÉRABILITÉ .....	<a href="#">43</a>
6.1	Portée de l'évaluation .....	<a href="#">43</a>
6.2	Choix des variables et collecte des données .....	<a href="#">43</a>
6.3	Ensemble des sujets à évaluer et présentation des données .....	<a href="#">44</a>
6.4	Évaluation statistique .....	<a href="#">45</a>
6.5	Sommaire intégré .....	<a href="#">46</a>
7.	RAPPORT .....	<a href="#">46</a>
7.1	Évaluation et rapport .....	<a href="#">46</a>
7.2	Sommaire de la base de données cliniques .....	<a href="#">49</a>
7.2.1	<i>Données relatives à l'efficacité</i> .....	<a href="#">49</a>
7.2.2	<i>Données relatives à l'innocuité</i> .....	<a href="#">50</a>
	GLOSSAIRE .....	<a href="#">51</a>

## 1. INTRODUCTION

### 1.1 Contexte et objet

On doit démontrer l'efficacité et l'innocuité des produits thérapeutiques au moyen d'essais cliniques menés selon la ligne directrice adoptée le 1<sup>er</sup> mai 1996 par l'ICH (*Les bonnes pratiques cliniques: directives consolidées - E6*). Dans cette ligne directrice de l'ICH on reconnaît comme essentiel le rôle de la conception et de l'analyse statistiques des essais cliniques. L'essor rapide de la recherche statistique dans le domaine des essais cliniques, combiné avec le rôle capital de la recherche clinique dans le processus d'homologation des médicaments et la prestation des soins de santé, oblige la publication d'un ouvrage succinct sur les questions statistiques ayant trait aux essais cliniques. La présente ligne directrice est rédigée principalement dans le but d'harmoniser les principes de méthodologie statistique appliqués aux essais cliniques des médicaments qui feront l'objet ultérieurement d'une demande de mise en marché en Europe, au Japon ou aux États-Unis.

Cette ligne directrice s'inspire du guide du CSP (*Comité des spécialités pharmaceutiques*) intitulé «Biostatistical Methodology in Clinical Trials in Applications for Marketing Authorisations for Medicinal Products» (décembre 1994), ainsi que de l'ouvrage du ministère japonais de la Santé et du Bien-être social, «Guidelines on the Statistical Analysis of Clinical Studies» (mars 1992), et de celui de la Food and Drug Administration des États-Unis, «Guideline for the Format and Content of the Clinical and Statistical Sections of a New Drug Application» (juillet 1988). Certaines questions touchant les principes et les méthodes statistiques se retrouvent dans d'autres lignes directrices de l'ICH, notamment celles mentionnées ci-dessous. Les lignes directrices qui ont rapport au contenu du présent document sont citées à l'endroit pertinent.

- E1A: Degré d'exposition de la population et évaluation de l'innocuité des médicaments
- E2A: Gestion des données sur l'innocuité clinique des médicaments: définitions et normes relatives à la déclaration rapide
- E2B: Gestion des données sur l'innocuité clinique des médicaments: éléments d'information à communiquer dans les rapports individuels sur les réactions indésirables et les incidents thérapeutiques liés à des médicaments
- E2C: Gestion des données sur l'innocuité clinique des médicaments: mise à jour périodique des rapports sur l'innocuité des médicaments commercialisés
- E3: Structure et contenu des rapports d'étude clinique
- E4: Données relatives à la relation dose-effet à l'appui de l'enregistrement des médicaments

- E5: Facteurs ethniques influant sur l'acceptabilité des données cliniques d'origine étrangère
- E6: Les bonnes pratiques cliniques: directives consolidées
- E7: Études à l'appui des groupes spéciaux: gériatrie
- E8: Considérations générales relatives aux études cliniques
- E10: Choice of Control Group in Clinical Trials
- M1: Standardisation of Medical Terminology for Regulatory Purposes
- M3: Calendrier des études d'innocuité non cliniques pour la conduite d'essais cliniques de produits pharmaceutiques.

Cette ligne directrice a pour but d'orienter les promoteurs pour la conception, la conduite, l'analyse et l'évaluation des essais cliniques de produits thérapeutiques dans le cadre plus global de la mise au point clinique. Elle guidera aussi les experts scientifiques qui sont chargés de rédiger les résumés de demande de mise en marché ou d'évaluer les données relatives à l'efficacité et à l'innocuité des médicaments tirées principalement d'essais cliniques avancés.

## 1.2 Champ d'application et orientation

Cette ligne directrice a pour point de mire les principes statistiques. Elle ne traite pas de l'utilisation de méthodes ou de procédés statistiques particuliers. Il revient plutôt au promoteur de voir à l'application conforme des principes statistiques. La ligne directrice traite par ailleurs de l'intégration des données de divers essais cliniques, mais cette question n'est pas un thème majeur dans l'ouvrage. L'application de principes et de procédés relatifs à la gestion des données ou au contrôle des essais cliniques fait l'objet d'autres lignes directrices de l'ICH et n'est pas examinée ici.

Cette ligne directrice devrait intéresser les membres d'un grand nombre de disciplines scientifiques. On suppose, toutefois, que l'aspect statistique des essais cliniques sera confié entièrement à la responsabilité d'un statisticien compétent et expérimenté, comme on le mentionne dans la ligne directrice E6 de l'ICH. Le rôle et la responsabilité du statisticien affecté à l'essai (voir le glossaire) seront de voir, en collaboration avec d'autres spécialistes des essais cliniques, à l'application conforme des principes statistiques dans les essais cliniques menés en vue de la mise au point de médicaments. Il devra donc posséder une formation et une expérience suffisamment grandes pour mettre en application les principes exposés dans cette ligne directrice.

Pour chaque essai clinique qui mène à une demande de mise en marché, on doit préciser tous les détails majeurs de la conception et de la conduite de l'essai, ainsi que les principales caractéristiques de l'analyse statistique proposée, dans un protocole rédigé avant que ne débute l'essai. Le degré de validité des résultats finals et des conclusions de l'essai dépendra



entre autres de la mesure dans laquelle on aura planifié *a priori* les étapes du protocole et l'analyse des données primaires. Le protocole, ainsi que les modifications éventuelles, devront être approuvés par le personnel compétent, y compris le statisticien. Celui-ci devra veiller à ce que le protocole, et toute modification éventuelle, traitent avec clarté et précision toutes les questions statistiques pertinentes en employant au besoin des termes techniques.

Les principes exposés dans cette ligne directrice concernent surtout les essais cliniques qui sont menés aux dernières phases de la mise au point, et bon nombre de ces essais visent essentiellement à confirmer l'efficacité du médicament. Outre l'efficacité, les essais confirmatoires peuvent avoir pour variable primaire une variable d'innocuité (p. ex., incident thérapeutique, variable de laboratoire clinique ou ECG), une variable pharmacodynamique ou encore une variable pharmacocinétique (comme dans les essais de bioéquivalence confirmatoires). De plus, on peut tirer des observations confirmatoires de données issues de divers essais à la fois et, de fait, la présente ligne directrice renferme des principes qui peuvent s'appliquer dans ce cas. Enfin, bien que les premières phases de la mise au point de médicaments consistent principalement dans des essais cliniques qui sont par définition exploratoires, les principes statistiques touchent aussi les essais de ce genre. Par conséquent, on devrait appliquer autant que possible le contenu du présent document à toutes les phases de la mise au point clinique.

Bon nombre des principes exposés dans la présente ligne directrice concernent les opérations qui visent à réduire au maximum le biais, ou erreur systématique, (voir le glossaire) et celles qui visent à porter au maximum le degré de précision. Au sens où il est employé dans cette ligne directrice, le terme «biais» désigne la tendance systématique de tout facteur lié à la conception, à la conduite et à l'analyse des essais cliniques, et à l'interprétation de leurs résultats, à créer un écart entre l'estimation de l'effet d'un traitement (voir le glossaire) et la valeur vraie correspondante. Il importe de définir le plus grand nombre possible de sources d'erreur systématique, de sorte que l'on puisse travailler à réduire au minimum ce type d'erreur. La présence d'erreurs systématiques peut compromettre sérieusement la validité des conclusions que l'on pourra tirer des essais cliniques. Le plan de l'essai peut être une source d'erreur systématique (par exemple, on pourrait décider d'un plan d'administration des traitements tel que les sujets à faible risque reçoivent systématiquement un seul et même traitement). La conduite et l'analyse des essais cliniques sont d'autres sources d'erreur systématique. Par exemple, le non-respect du protocole et le fait d'exclure des sujets de l'analyse sous prétexte que l'on connaît déjà les résultats du traitement qui leur serait administré sont des sources d'erreur systématique qui peuvent empêcher d'évaluer avec précision l'effet du traitement.

Comme le biais peut être introduit d'une manière subtile ou inconnue et qu'on ne peut en mesurer directement l'effet, il est important d'évaluer la robustesse des résultats et des conclusions préliminaires de l'essai. La robustesse est le concept selon lequel les conclusions générales d'un test sont plus ou moins sensibles à divers aspects des données, des hypothèses et des méthodes d'analyse des données. La robustesse signifie que l'effet du traitement et les conclusions préliminaires de l'essai ne seront pas modifiés sensiblement si l'analyse est exécutée selon d'autres hypothèses ou avec d'autres méthodes que celles utilisées initialement. L'interprétation des mesures statistiques de l'incertitude de l'effet du traitement et des comparaisons de traitements devrait prendre en compte l'incidence probable de l'erreur systématique sur la valeur  $p$ , l'intervalle de confiance ou l'inférence.

Étant donné que la conception et l'analyse des essais cliniques font intervenir principalement des méthodes statistiques axées sur l'observation de fréquences (voir le glossaire), cette ligne directrice traite abondamment de l'utilisation de ce genre de méthodes lorsqu'il est question de test d'hypothèses ou d'intervalle de confiance. Il ne faudrait pas croire pour autant que les autres méthodes ne sont pas convenables: on peut en effet envisager l'utilisation de méthodes bayésiennes (voir le glossaire) ou d'autres types de méthodes lorsque les raisons de leur utilisation sont claires et que les conclusions qui en découlent sont suffisamment robustes.

## 2. CONSIDÉRATIONS RELATIVES À LA MISE AU POINT CLINIQUE DES MÉDICAMENTS

### 2.1 Contexte des essais

#### 2.1.1 Plan de mise au point

L'objectif général de la mise au point clinique d'un nouveau médicament est de déterminer la plage de doses et le régime qui font qu'un médicament est à la fois sûr et efficace, sous réserve d'un rapport risques-avantages acceptable. Par ailleurs, il faut définir les sujets auxquels le médicament peut être bénéfique, ainsi que les indications et la posologie du médicament.

Pour atteindre ces objectifs, il faut un programme ordonné d'essais cliniques, avec des objectifs propres à chaque essai (voir ligne directrice E8 de l'ICH). Ces renseignements doivent être consignés dans un plan de mise au point clinique ou une série de plans, avec les points de décision pertinents et la souplesse nécessaire pour pouvoir apporter des modifications au fil de l'accumulation des connaissances. La demande de mise en marché doit décrire clairement l'essentiel du contenu de ces plans et la contribution de chaque essai. L'interprétation et l'évaluation des résultats du

programme d'essais impliquent que l'on fasse une synthèse des résultats de chacun des essais (voir section 7.2). Pour faciliter cette synthèse, on veille à ce que des normes communes soient adoptées pour certains aspects des essais (p. ex., dictionnaires de termes médicaux, définition des principales mesures et calendrier d'observation, correction des écarts au protocole, et ainsi de suite). Lorsque l'examen de questions médicales passe par plus d'un essai, il peut être utile de produire un résumé statistique ou un aperçu de la question ou d'effectuer une méta-analyse (voir le glossaire). On devrait, si possible, inclure ces opérations dans le plan, de sorte que les essais pertinents soient clairement définis et les caractéristiques communes nécessaires des plans correspondants, définies à l'avance. Enfin, le plan en question doit prendre en compte les autres grandes questions statistiques (s'il en est) qui sont censées influencer sur certains des essais.

### **2.1.2 Essai confirmatoire**

L'essai confirmatoire est un essai adéquatement contrôlé dans lequel les hypothèses sont énoncées à l'avance, puis évaluées. En principe, l'essai confirmatoire est nécessaire pour démontrer clairement l'efficacité ou l'innocuité d'un médicament. Dans ce genre d'essai, la principale hypothèse étudiée découle directement de l'objectif premier de l'essai, elle est toujours définie à l'avance et elle est celle qui est testée une fois que l'essai est terminé. Il est également important, dans un essai confirmatoire, d'estimer avec précision les effets du traitement étudié et d'établir un lien entre ces effets et leur importance clinique.

Les essais confirmatoires ont pour but de vérifier d'une manière non équivoque la validité de revendications; le respect du protocole et des modes opératoires normalisés est donc particulièrement important. On doit expliquer et justifier tout changement indispensable et on doit en analyser l'effet. Par ailleurs, le protocole doit contenir la justification du plan de chaque essai et celle d'autres aspects statistiques importants, comme les caractéristiques principales de l'analyse prévue. On doit limiter le nombre de questions sur lesquelles l'essai doit porter.

Pour que l'on puisse vérifier d'une manière non équivoque la validité des revendications, il faut que les résultats des essais confirmatoires démontrent que le produit testé présente des avantages cliniques. L'essai confirmatoire devrait donc suffire pour répondre d'une façon claire et définitive à chaque question clinique touchant une allégation relative à l'efficacité ou à l'innocuité. En outre, il est important de bien comprendre, et de bien faire saisir, ce sur quoi repose la généralisation (voir le glossaire); cela peut par ailleurs influencer sur le nombre et le type (p. ex., spécialiste ou

généraliste) de centres ou d'essais requis. L'essai confirmatoire doit produire des résultats robustes. Dans certaines circonstances, le poids de la preuve d'un seul essai confirmatoire pourra suffire.

### **2.1.3 Essai exploratoire**

La justification et le plan d'un essai confirmatoire reposent presque toujours sur des recherches cliniques antérieures menées dans le cadre d'une série d'études exploratoires. Comme tous les essais cliniques, ces études doivent avoir des objectifs clairs et précis, mais contrairement aux objectifs des essais confirmatoires, ceux des études exploratoires ne mènent pas toujours à des tests d'hypothèses définies au préalable. En outre, les essais exploratoires peuvent parfois exiger un plan plus souple, de sorte que l'on puisse apporter des changements au fil de l'accumulation des résultats. L'analyse des essais exploratoires peut nécessiter l'exploration de données; on peut exécuter des tests d'hypothèse, mais le choix de l'hypothèse pourra dépendre des données. Enfin, la démonstration formelle de l'efficacité d'un médicament ne peut reposer sur des essais exploratoires, quoique ceux-ci puissent contribuer à la constitution du dossier.

On peut retrouver des aspects de chaque type d'essai dans un seul et même essai. Par exemple, dans la plupart des essais confirmatoires, les données sont soumises à des analyses exploratoires qui servent à expliquer ou à confirmer les résultats des essais et à proposer d'autres hypothèses pour la recherche future. Le protocole doit établir une distinction claire entre les aspects de l'essai qui serviront à l'épreuve de confirmation et ceux qui serviront à l'analyse exploratoire des données.

## **2.2 Champ d'application des essais**

### **2.2.1 Population**

Dans les premières phases de la mise au point d'un médicament, le choix des sujets qui doivent participer à un essai clinique peut être fortement influencé par le désir de porter au maximum la probabilité d'observer les effets cliniques qui font l'objet de la recherche; il se peut donc que les sujets proviennent d'un sous-groupe très restreint de la population des patients auxquels pourrait être éventuellement destiné le médicament. Cependant, au moment des essais confirmatoires, les sujets devraient être plus représentatifs de la population cible. C'est pourquoi il est généralement utile, dans ces essais, d'assouplir autant que possible les critères d'inclusion et d'exclusion par rapport à la population cible, tout en conservant une homogénéité suffisamment grande pour

pouvoir estimer avec précision les effets du traitement. On ne saurait s'attendre à ce qu'un essai clinique soit parfaitement représentatif des cas d'utilisation futurs à cause de l'effet probable du lieu géographique, du moment où l'essai est mené, des pratiques médicales des chercheurs ou de la clinique, et ainsi de suite. Néanmoins, on devrait réduire dans toute la mesure du possible l'effet de ces facteurs et examiner cet effet au moment de l'interprétation des résultats de l'essai.

### **2.2.2 Variables primaires et secondaires**

La variable primaire (variable «cible», critère primaire) est la variable capable de fournir l'information la plus pertinente et la plus probante cliniquement qui se rapporte directement à l'objectif premier de l'essai. Normalement, il ne doit y avoir qu'une seule variable primaire; ce sera le plus souvent une variable d'efficacité, car l'objectif premier de la plupart des essais confirmatoires est de démontrer scientifiquement de façon probante l'efficacité d'un médicament. La variable primaire pourra parfois représenter l'innocuité ou la tolérance, qui seront toujours des critères importants. On pourra aussi avoir comme variable primaire des mesures ayant trait à la qualité de vie ou à l'économie de la santé. Le choix de la variable primaire doit être en conformité avec les normes reconnues du domaine de recherche pertinent. Il est recommandé d'utiliser une variable fiable et validée, qui a déjà fait l'objet d'études antérieures ou de publications. On doit être suffisamment sûr que la variable primaire choisie produira une mesure fiable et valide de l'existence d'un effet clinique favorable pertinent et notable pour la population de patients définie par les critères d'inclusion et d'exclusion. En règle générale, la variable primaire doit être la variable utilisée pour estimer la taille de l'échantillon (voir section 3.5).

Dans beaucoup de cas, la méthode d'évaluation des effets du traitement peut être complexe, et elle doit donc être décrite soigneusement. Par exemple, il est inacceptable de spécifier la mortalité comme une variable primaire sans aucune clarification; en effet, on peut évaluer la mortalité soit en comparant la proportion de sujets vivants à des périodes déterminées ou en comparant les distributions globales des durées de survie pour un intervalle donné. Un autre exemple courant est le cas de l'événement périodique; là encore, la mesure de l'effet du traitement peut consister en une simple variable dichotomique (apparition d'un événement durant une période donnée), ou elle peut être la période écoulée avant le premier événement, la fréquence d'apparition (nombre d'événements par période pour une unité d'observation), etc. Dans l'étude des traitements pour les maladies chroniques, l'évaluation de la capacité fonctionnelle à diverses périodes pose d'autres difficultés pour le choix de la variable primaire. En effet, les méthodes d'évaluation sont nombreuses: comparaison des évaluations faites au

début et à la fin de la période d'observation, comparaison des valeurs de la pente calculées dans chaque évaluation pour toute la période, comparaison des proportions de sujets qui se trouvent au-dessus ou au-dessous d'un seuil déterminé, ou comparaison fondée sur des méthodes pour mesures répétées. Pour éviter les problèmes de multiplicité qui découlent de définitions *post hoc*, il est absolument nécessaire de définir avec exactitude dans le protocole la variable primaire qui sera utilisée dans l'analyse statistique. En outre, le protocole devra normalement traiter de la pertinence clinique de la variable primaire choisie et de la validité des méthodes de mesure pertinentes et en présenter la justification.

Le protocole doit mentionner la variable primaire, ainsi que les raisons de ce choix. On pourra très rarement redéfinir la variable primaire après la levée de l'insu, puisque la redéfinition introduit des erreurs systématiques difficiles à évaluer. Lorsqu'il faut mesurer de plusieurs façons l'effet clinique défini par l'objectif primaire, le protocole doit définir l'une des mesures comme la variable primaire - suivant la pertinence clinique de ces mesures, leur importance, leur objectivité et d'autres caractéristiques pertinentes - quand il est possible de faire un tel choix.

Les variables secondaires, quant à elles, sont soit des mesures d'appoint liées à l'objectif primaire ou des mesures d'effets liées aux objectifs secondaires. Il est tout aussi important de les définir au préalable dans le protocole et d'en expliquer l'importance relative et le rôle dans l'interprétation des résultats de l'essai. On doit restreindre le nombre des variables secondaires, comme on l'a fait pour les questions auxquelles tente de répondre l'essai.

### 2.2.3 Variables composées

Si l'on ne peut choisir une variable primaire parmi les nombreuses mesures liées à l'objectif premier, on gagnera à intégrer ou à combiner les diverses mesures dans une seule variable, appelée variable «composée», au moyen d'un algorithme défini au préalable. De fait, la variable primaire se présente parfois comme la combinaison de plusieurs mesures cliniques (p. ex., les échelles d'évaluation utilisées dans l'étude de l'arthrite, des troubles psychiatriques et d'autres maladies). Cette méthode permet de résoudre le problème de la multiplicité sans qu'il soit nécessaire d'apporter des ajustements à l'erreur de première espèce. Le protocole doit mentionner l'utilisation de cette méthode, le cas échéant, et il doit indiquer comment interpréter l'échelle qui résultera de cette opération par rapport à ce qui pourrait être considéré comme un effet clinique favorable pertinent.

Lorsqu'une variable composée sert de variable primaire, on peut parfois analyser séparément chacune des composantes, si celles-ci sont significatives et valides sur le plan clinique. Lorsque c'est une échelle d'évaluation qui sert de variable primaire, il est particulièrement important de considérer des facteurs tels que la validité de contenu (voir le glossaire), la fiabilité inter-évaluateur et intra-évaluateur (voir le glossaire) et la rapidité de détection de changements dans la gravité d'une maladie.

#### **2.2.4 Variables d'évaluation globale**

Dans certains cas, on définit des variables «d'évaluation globale» (voir le glossaire) afin de mesurer l'innocuité, l'efficacité et l'utilité globales d'un traitement. Ce type de variable intègre des critères objectifs et l'opinion générale du chercheur concernant l'état du sujet ou l'évolution de cet état, et elle consiste normalement dans une échelle de cotes nominales ordonnées. L'évaluation globale de l'efficacité est une pratique bien établie dans certains domaines thérapeutiques comme la neurologie et la psychiatrie.

Les variables d'évaluation globale renferment ordinairement une composante subjective. Lorsqu'une variable d'évaluation globale sert de variable primaire ou secondaire, le protocole doit contenir des détails précis sur l'échelle, à savoir:

- 1) le rapport entre l'échelle et l'objectif premier de l'essai;
- 2) le fondement de la validité et de la fiabilité de l'échelle;
- 3) comment se servir des données recueillies sur un sujet pour classer celui-ci dans une seule et unique catégorie;
- 4) comment classer des sujets au dossier incomplet dans une seule et unique catégorie ou, à défaut, comment évaluer ces cas.

Si le chercheur considère des variables objectives dans une évaluation globale, ces variables doivent être définies comme des variables primaires additionnelles ou, à tout le moins, comme des variables secondaires majeures.

L'évaluation globale de l'utilité d'un médicament prend en compte les risques aussi bien que les avantages et elle reproduit le processus décisionnel que suit le médecin traitant, qui doit peser le pour et le contre dans ses décisions concernant l'utilisation d'un produit. Un des inconvénients des variables d'utilité globale est que leur utilisation peut parfois avoir pour conséquence de faire reconnaître deux produits comme équivalents,

même si leurs bilans des effets favorables et indésirables sont très différents. Par exemple, le fait de considérer qu'un traitement a une utilité globale équivalente ou supérieure à celle d'un autre traitement, parce qu'il a moins d'effets indésirables, peut occulter le fait que le premier traitement a une efficacité à peu près nulle. C'est pourquoi il n'est pas recommandé d'utiliser la variable d'utilité globale comme une variable primaire. Mais si jamais elle l'était, il serait important de définir les variables d'efficacité et d'innocuité comme autant de variables primaires additionnelles.

### **2.2.5 Variables primaires multiples**

Il est parfois souhaitable d'utiliser plus d'une variable primaire pour représenter toute la gamme des effets des traitements. Il conviendra alors d'exposer clairement le mode d'interprétation qu'on prévoit utiliser pour ce type d'information. De même, on devra indiquer clairement si l'on juge nécessaire qu'une seule variable, qu'un nombre minimum de variables ou que toutes les variables soient touchées pour que les objectifs de l'essai soient atteints. On devra énoncer clairement l'hypothèse (ou les hypothèses) primaire(s), ainsi que les paramètres étudiés (p. ex., moyenne, pourcentage, distribution), par rapport aux variables primaires définies, et on devra décrire la méthode d'inférence statistique. On devra par ailleurs expliquer l'incidence sur l'erreur de première espèce, à cause de l'apparition possible de problèmes de multiplicité (voir section 5.6); le protocole devra aussi décrire la méthode utilisée pour limiter l'erreur de première espèce. On peut examiner le degré d'intercorrélation entre les variables primaires proposées en évaluant l'incidence sur l'erreur de première espèce. Si l'essai a pour but de montrer les effets qui s'exercent sur toutes les variables primaires définies, il n'est pas nécessaire d'apporter des ajustements à l'erreur de première espèce, mais il convient d'examiner soigneusement l'incidence sur l'erreur de seconde espèce et la taille de l'échantillon.

### **2.2.6 Variables de substitution**

Lorsqu'il n'est pas possible d'évaluer directement l'effet clinique favorable d'un traitement pour un sujet par l'observation de l'efficacité clinique, on peut envisager l'utilisation de critères indirects (variables de substitution - voir le glossaire). On utilise des variables de substitution généralement reconnues dans un certain nombre d'indications où elles devraient être des prédicteurs fiables de l'effet clinique favorable. L'introduction d'une variable de substitution soulève deux grandes questions. Premièrement, elle pourrait ne pas être un bon prédicteur de l'effet clinique recherché. Par exemple, elle pourrait mesurer l'effet du traitement lié à un mécanisme pharmacologique particulier, mais ne pas fournir toute l'information voulue sur la



gamme des activités et les effets ultimes du traitement, qu'ils soient positifs ou négatifs. Nombreux sont les cas où des traitements qui avaient eu un effet clinique très positif sur une variable de substitution se sont avérés néfastes en définitive pour les sujets; à l'inverse, il y a des traitements pour lesquels on a observé un effet clinique favorable sans avoir pu mesurer leur incidence sur des variables de substitution. Deuxièmement, la variable de substitution pourrait ne pas produire de mesure quantitative de l'effet clinique favorable qui puisse être mise en balance directement avec les effets indésirables. On a proposé des critères statistiques pour la validation des variables de substitution, mais on a encore peu d'expérience dans leur utilisation. Dans la pratique, la qualité de l'information obtenue par la variable de substitution dépendra de trois facteurs: 1) la plausibilité biologique de la relation, 2) les études épidémiologiques expérimentales de la capacité de prédiction de la variable de substitution pour le résultat clinique, et 3) les résultats d'essais cliniques qui montrent que le traitement a des effets correspondants sur la variable de substitution et sur le résultat clinique. La relation qui existe entre le résultat clinique et la variable de substitution pour un produit donné n'est pas nécessairement observé pour un autre produit destiné à traiter la même maladie, mais agissant d'une manière différente.

### **2.2.7 Variables qualitatives catégorielles**

Il est parfois souhaitable de dichotomiser ou d'assigner d'une manière quelconque des variables continues ou ordinales. Les critères «succès» et «réponse» sont des exemples courants de dichotomie, où l'on doit, par exemple, fixer un pourcentage d'amélioration minimum (seuil relatif) pour une variable continue ou définir un classement selon lequel les valeurs d'une échelle d'évaluation ordinale sont égales ou supérieures à un seuil donné (p. ex., «bon»). La réduction de la tension artérielle diastolique sous 90 mmHg est un exemple typique de dichotomisation. La catégorisation est surtout utile lorsque sa pertinence sur le plan clinique est démontrée. On doit définir au préalable les critères de classement et les mentionner dans le protocole, car la connaissance des résultats d'un essai pourrait facilement influencer le choix de ces critères. Comme la catégorisation implique normalement une perte d'information, l'analyse s'en trouve affaiblie; on doit en tenir compte dans la détermination de la taille de l'échantillon.

## 2.3 Méthodes de réduction du biais

L'insu et la randomisation sont les principales méthodes de réduction du biais utilisées dans les essais cliniques, et elles devraient faire partie de la plupart des essais cliniques contrôlés destinés à appuyer une demande de mise en marché. La majorité des essais de ce genre sont des épreuves à double insu, où les traitements sont pré-emballés suivant une formule de randomisation appropriée, puis remis à la clinique responsable des essais avec une étiquette sur laquelle figurent uniquement le numéro du sujet et la période de traitement, de sorte qu'aucun des intervenants ne connaisse le traitement administré au sujet, pas même sous forme de code alphabétique. Cette façon de procéder est celle proposée dans la section 2.3.1 et dans la majeure partie de la section 2.3.2, à quelques exceptions près qui seront mentionnées à la fin.

On peut aussi amenuiser les risques de biais au moment de l'élaboration du plan en inscrivant dans le protocole des modalités qui visent à réduire au minimum le nombre d'irrégularités qui pourraient se produire dans la conduite de l'essai et nuire à l'analyse (p. ex., violations du protocole, retrait de sujets et valeurs manquantes). Le protocole devrait aussi proposer des façons de réduire la fréquence de ces problèmes et de résoudre les difficultés qui surviennent dans l'analyse des données.

### 2.3.1 *Insu*

L'insu vise à limiter l'introduction d'un biais volontaire ou involontaire dans la conduite des essais cliniques et l'interprétation des résultats; ce biais découlerait de l'influence que peut avoir la connaissance du traitement sur le recrutement et l'affectation des sujets, les soins à leur donner, l'attitude des sujets à l'égard des traitements, l'évaluation des paramètres, le traitement des cas de retrait, l'exclusion de données de l'analyse, et ainsi de suite. L'objectif essentiel de l'insu est d'empêcher l'identification des traitements jusqu'à ce que tous les risques d'erreur systématique soient levés.

L'essai à double insu est une épreuve où ni le sujet, ni le personnel du chercheur ou du promoteur affecté au traitement ou à l'évaluation clinique du sujet ne connaissent le traitement administré. Cela comprend les personnes chargées de déterminer l'admissibilité du sujet, de mesurer les paramètres ou d'évaluer la conformité au protocole. L'insu est maintenu pendant toute la durée de l'essai et il n'est levé pour le personnel concerné que lorsque les données affichent un niveau de qualité acceptable après traitement. Si l'on devait permettre à des membres du personnel du promoteur qui ne sont pas affectés au traitement ou à l'évaluation clinique des sujets de connaître les codes du traitement (p. ex., spécialistes de l'analyse biologique, vérificateurs, personnes chargées de rapporter les événements indésirables sérieux), le promoteur

devra prévenir la diffusion non autorisée des codes de traitement par l'application de modes opératoires normalisés appropriés. Dans un essai à simple insu, le chercheur (et/ou son personnel) est au courant du traitement, mais non le sujet, ou vice versa. Dans un essai ouvert, l'identité du traitement est connue de tous. L'essai à double insu est la formule optimale; en effet, il n'est pas possible dans ce cas de distinguer (apparence, goût, etc), ni avant ni pendant, les traitements qui doivent être administrés durant l'essai, et l'insu doit être maintenu durant tout l'essai.

En contrepartie, la réalisation d'un essai à double insu n'est pas exempte de difficultés: les traitements peuvent être de nature totalement différente, par exemple chirurgie vs pharmacothérapie; deux médicaments peuvent avoir des formulations différentes et, bien que l'utilisation de capsules les rende indistinguables, le changement de formulation pourrait modifier les propriétés pharmacocinétiques ou pharmacodynamiques et obliger par conséquent le chercheur à établir la bioéquivalence des formulations; deux traitements peuvent avoir des régimes d'administration quotidiens différents. Une façon de réaliser les conditions d'un essai à double insu dans ces circonstances est d'utiliser la technique de double placebo (voir le glossaire). Cette technique peut parfois imposer un régime suffisamment anormal pour nuire à la motivation du sujet et le décourager de se conformer à l'essai. Par ailleurs, son utilisation peut être freinée par des questions d'éthique, par exemple lorsqu'elle implique le recours à des modes opératoires factices. Néanmoins, on devrait faire tous les efforts nécessaires pour résoudre ces difficultés.

Les effets apparents d'un traitement chez des sujets peuvent compromettre en partie le caractère «double insu» de certains essais cliniques. Dans ces circonstances, on peut renforcer l'insu en cachant aux chercheurs et aux membres du personnel du promoteur certains résultats d'essai (p. ex., données de laboratoire particulières). Par ailleurs, on devrait envisager des méthodes semblables pour réduire le biais (voir ci-dessous) dans des essais ouverts où l'apparition d'effets de traitements particuliers peut mener au décodage de l'insu à l'égard d'un sujet.

Si on ne peut mener un essai à double insu, on doit envisager l'essai à simple insu. Dans certains cas, seul l'essai ouvert est envisageable du point de vue pratique ou du point de vue éthique. L'essai à simple insu et l'essai ouvert offrent plus de souplesse, mais il est essentiel que la connaissance que peut avoir le chercheur du traitement suivant n'influe pas sur la décision d'inscrire ou non le sujet; le chercheur devrait prendre cette décision avant de connaître le traitement randomisé. Pour ces essais, on devrait envisager l'utilisation d'une méthode de randomisation centralisée, comme la randomisation par téléphone, pour administrer les affectations de traitements randomisés. En outre, les évaluations cliniques devraient être faites par des membres du personnel médical qui ne

sont pas chargés de traiter les sujets et qui ne savent rien du traitement administré. On ne devrait ménager aucun effort pour réduire au minimum les diverses sources d'erreur systématique connues dans les essais à simple insu et les essais ouverts, et les variables primaires devraient être aussi objectives que possible. Dans le protocole, on devrait expliquer pourquoi on a opté pour tel degré d'insu et on devrait exposer les autres mesures prises pour réduire au minimum les risques de biais. Par exemple, le promoteur doit avoir prévu des modes opératoires normalisés convenables pour limiter l'accès aux codes des traitements durant l'épuration de la base de données qui précède l'analyse.

On ne peut envisager de décoder l'insu à l'égard d'un sujet que lorsque le médecin traitant juge qu'il est essentiel de connaître l'affectation du traitement pour le bien du sujet. À la fin d'un essai où s'est produit un décodage, intentionnel ou non, de l'insu, on doit faire rapport et donner les raisons de cet incident, quelles qu'elles soient. Enfin, on doit expliquer dans le protocole comment, et à quel moment, on révélera les affectations de traitements.

Dans le présent document, l'examen aveugle des données (voir le glossaire) désigne l'opération de vérification des données qui se déroule entre le moment où l'essai prend fin (dernière observation sur le dernier sujet) et le moment où l'insu est levé.

### **2.3.2 Randomisation**

La randomisation introduit expressément un élément de hasard dans le processus d'affectation des traitements durant un essai clinique. À l'étape de l'analyse des données, la randomisation offre un cadre statistique objectif pour l'évaluation quantitative de l'information relative aux effets du traitement. Elle tend aussi à produire des groupes de traitement pour lesquels les distributions des facteurs pronostiques, connus et inconnus, sont semblables. Avec l'insu, la randomisation permet d'éliminer les biais de sélection et de répartition des sujets qui découlent de la prévisibilité des affectations de traitements.

Le plan de randomisation d'un essai clinique décrit l'affectation aléatoire des traitements entre les sujets. Dans sa forme la plus élémentaire, ce plan consiste dans une liste séquentielle des traitements (ou des séries de traitements, dans le cas d'un essai croisé), ou des codes correspondants, par numéro de sujet. L'organisation de certains essais, comme ceux qui comportent une phase de présélection, peut compliquer la situation, mais le mode d'affectation des traitements ou des séries de traitements, qui est

établi à l'avance, devrait être clair. Le mode d'élaboration des plans de randomisation variera selon le plan de l'essai. Le plan de randomisation doit être reproductible (si besoin est).

Bien que la randomisation sans restriction soit une méthode acceptable, la randomisation par blocs comporte généralement des avantages. En effet, cette dernière méthode accroît la comparabilité des groupes de traitement, surtout lorsque les caractéristiques des sujets peuvent changer au fil du temps à cause, par exemple, de la modification de la politique de recrutement. De plus, cette méthode permet plus facilement d'obtenir des groupes de traitement de taille équivalente. Dans des essais croisés, la randomisation par blocs permet d'obtenir des plans équilibrés, qui sont par définition plus efficaces et plus faciles à interpréter. On veillera à choisir des blocs suffisamment petits pour limiter les risques de déséquilibre, mais aussi suffisamment grands pour empêcher toute prédiction lorsque la fin d'une séquence approche dans un bloc. Les investigateurs et les autres membres du personnel concernés ne doivent pas connaître la dimension du bloc; l'utilisation de deux dimensions ou plus, déterminées aléatoirement pour chaque bloc, est un autre moyen de dissimuler la dimension. (Théoriquement, la prévisibilité importe peu dans un essai à double insu, mais les effets pharmacologiques du médicament peuvent amener des investigateurs à élaborer des hypothèses astucieuses.)

Dans des essais multicentriques (voir le glossaire), la randomisation doit être organisée de manière centrale. Il est recommandé d'avoir un plan pour chaque centre, c'est-à-dire de faire une stratification pour chaque centre ou d'affecter plusieurs blocs entiers à chaque centre. D'une manière plus générale, il est parfois utile d'effectuer une stratification selon les facteurs pronostiques fondamentaux (p. ex., gravité de la maladie, âge, sexe, etc.) dans le but de réaliser une répartition équilibrée à l'intérieur des strates; cette méthode est encore plus profitable pour les essais de taille modeste. L'utilisation de plus de deux ou trois facteurs de stratification est rarement nécessaire; en outre, cette formule est moins efficace pour réaliser l'équilibre et son application est contraignante. L'utilisation d'une méthode de répartition dynamique (voir ci-dessous) peut contribuer à réaliser l'équilibre entre un certain nombre de facteurs de stratification, pourvu que l'on puisse adapter les autres procédures en conséquence. Les facteurs de stratification qui ont servi dans la randomisation doivent être pris en compte dans les étapes ultérieures de l'analyse.

Dans un essai randomisé, on doit toujours affecter à un nouveau sujet le traitement qui correspond au premier numéro disponible dans le plan de randomisation pertinent (ou dans la strate pertinente, si la randomisation est stratifiée). L'affectation ne doit se faire que lorsque est confirmée l'entrée du sujet dans la portion randomisation de l'essai. Le protocole d'essai ne doit pas faire état des caractéristiques de la randomisation qui favoriseraient la prévisibilité (p. ex., dimension des blocs). Le promoteur, ou une partie indépendante, verra à classer en toute sécurité le plan de randomisation, de manière que l'insu soit maintenu pendant toute la durée de l'essai. Le protocole d'accès au plan de randomisation devra prévoir que, dans une situation d'urgence, l'insu pourra être levé pour un sujet donné durant l'essai. La procédure à suivre, la documentation requise, de même que le traitement affecté au sujet par la suite et l'évaluation qu'on en fera, doivent être mentionnés dans le protocole de l'essai.

L'affectation dynamique est une méthode où les affectations de traitements sont déterminées par la situation d'équilibre courante des traitements déjà affectés ou, s'il s'agit d'un essai stratifié, par la strate à laquelle appartient le sujet et la situation d'équilibre dans cette strate. On doit éviter d'utiliser des méthodes d'affectation dynamique déterministes; on doit introduire plutôt un facteur de randomisation approprié pour chaque affectation de traitements. Il faut tout mettre en oeuvre pour conserver le caractère «à double insu» de l'essai. Par exemple, on peut faire en sorte que le code du traitement ne soit connu qu'à un bureau central, d'où est contrôlée l'affectation dynamique, le plus souvent par liaison téléphonique, ce qui permet de vérifier de nouveau les critères d'admissibilité et de décider de la participation d'un individu à l'essai, opérations qui peuvent être utiles pour certains types d'essai multicentrique. On peut ensuite appliquer le mode habituel de pré-emballage et d'étiquetage des médicaments pour les essais à double insu, mais ces médicaments ne sont plus utilisés dans un ordre séquentiel. Il est souhaitable, par ailleurs, d'utiliser des algorithmes pour faire en sorte que le personnel du bureau central ne connaisse pas les codes de traitement. Lorsqu'on envisage l'affectation dynamique, on doit mesurer soigneusement la complexité de la démarche et son effet probable sur l'analyse.

### 3. CONSIDÉRATIONS RELATIVES AU PLAN DE L'ESSAI

#### 3.1 Configuration du plan

##### 3.1.1 Plan à groupes parallèles

Le plan le plus couramment utilisé pour les essais confirmatoires est le plan à groupes parallèles, selon lequel les sujets sont répartis par randomisation entre deux groupes d'essai ou plus, chaque groupe étant associé à un traitement. Les traitements comprennent l'administration du produit étudié à une ou plusieurs doses, plus un ou plusieurs traitements témoins (p. ex., placebo, comparateur actif). Ce plan repose sur des hypothèses moins complexes que celles qui se rattachent à la plupart des autres plans. Toutefois, comme pour les autres plans, l'essai peut présenter des caractéristiques nouvelles qui viennent compliquer l'analyse et l'interprétation des résultats (p. ex., covariables, mesures répétées, interaction entre les facteurs du plan, écarts au protocole, abandons (voir le glossaire) et retraits).

##### 3.1.2 Plan croisé

Dans un plan croisé, on affecte par randomisation une série de deux traitements ou plus à chaque sujet; le sujet est donc son propre témoin pour les comparaisons de traitements. Cette méthode simple est intéressante surtout parce qu'elle réduit, parfois largement, le nombre des sujets et, généralement, le nombre des évaluations nécessaires pour atteindre une puissance donnée. Dans la version la plus élémentaire du plan croisé (2 x 2), chaque sujet reçoit deux traitements dans deux périodes successives, souvent séparées par une période de sevrage; l'affectation se fait par randomisation. L'autre version la plus courante du plan croisé implique la comparaison de  $n$  ( $>2$ ) traitements dans  $n$  périodes, chaque sujet recevant tous les traitements. Il existe de nombreuses variantes, par exemple le plan selon lequel chaque sujet reçoit un sous-ensemble des  $n$  ( $>2$ ) traitements, ou celui selon lequel le sujet reçoit plus d'une fois le même traitement.

Le plan croisé présente un certain nombre de faiblesses qui peuvent invalider ses résultats. La principale difficulté concerne l'effet différé, c'est-à-dire l'effet résiduel d'un traitement dans les périodes de traitement suivantes. Dans un modèle additif, l'existence d'un effet rémanent inégal aura pour conséquence d'introduire un biais dans les comparaisons directes de traitements. Dans un plan 2 x 2, on ne peut distinguer statistiquement l'effet rémanent de l'interaction entre le traitement et la période, et le

test applicable à l'un et l'autre de ces effets n'est pas suffisamment puissant car la comparaison correspondante se fait entre le «sujet et lui-même». Ce problème est moins aigu dans les plans d'ordre supérieur, mais on ne peut en faire abstraction totalement.

Il est donc important d'écarter l'effet rémanent lorsqu'on a recours au plan croisé. La meilleure façon de le faire sera par un usage sélectif et soigneux du plan, soutenu par une bonne connaissance du type de maladie et du nouveau médicament. La maladie à l'étude doit être une maladie chronique qui évolue peu. Les effets utiles du médicament doivent se manifester entièrement durant la période de traitement. Les périodes de sevrage doivent être suffisamment longues pour que l'organisme ait le temps d'éliminer complètement le médicament. On devra établir au préalable si ces conditions peuvent être respectées en se servant de données et d'informations *a priori*.

Les essais croisés comportent d'autres problèmes qui méritent une attention spéciale. Le problème le plus notable a trait aux complications qui découlent de la perte de sujets pour l'analyse et l'interprétation. En outre, le risque d'effet rémanent complique la tâche de rattacher au traitement pertinent les événements indésirables qui surviennent dans les périodes de traitement ultérieures. Ces problèmes, entre autres, sont décrits dans la ligne directrice E4 de l'ICH. En règle générale, on ne doit utiliser le plan croisé que dans les cas où l'on prévoit des pertes minimales de sujets.

Le plan croisé 2 x 2 sera souvent utilisé, généralement de façon satisfaisante, pour démontrer la bioéquivalence de deux formulations du même médicament. Lorsqu'on teste cette bioéquivalence chez des volontaires en santé, il est très peu probable que l'on puisse observer l'incidence de l'effet rémanent sur la variable pharmacocinétique pertinente si la période de sevrage est suffisamment longue. Néanmoins, il est important de tester cette hypothèse durant l'analyse sur la foi des données obtenues, par exemple en démontrant que le médicament n'est pas décelé au début d'une période.

### 3.1.3 Plans factoriels

Un plan factoriel sert à évaluer simultanément deux traitements ou plus par diverses combinaisons de ces traitements. Le cas le plus simple est le plan factoriel 2 x 2, où les sujets sont répartis aléatoirement entre quatre combinaisons possibles de traitements, qui sont au nombre de 2 (disons A et B), à savoir A seul, B seul, A et B, ni A ni B. Dans beaucoup de cas, on utilisera ce plan expressément pour étudier l'interaction de A et B. Le test statistique pertinent peut ne pas être assez puissant pour déceler une



interaction si la taille de l'échantillon a été déterminée en fonction du test des effets majeurs. Cette considération est importante lorsqu'on se sert de ce plan pour étudier les effets combinés de A et de B en particulier, si les traitements risquent d'être appliqués ensemble.

Une autre application importante du plan factoriel consiste à définir les caractéristiques dose-effet de l'utilisation simultanée des traitements C et D, en particulier lorsque l'efficacité de chaque monothérapie a été démontrée dans des essais antérieurs pour une dose donnée. On choisit un nombre  $m$  de doses de C, y compris généralement une dose zéro (placebo), et un nombre comparable,  $n$ , de doses de D. Le plan consiste alors dans  $m \times n$  groupes de traitement, chacun recevant une combinaison différente de doses de C et de D. On peut ensuite se servir de l'estimation de la surface de réponse obtenue pour définir une combinaison appropriée des doses de C et de D pour l'usage clinique (voir la ligne directrice E4 de l'ICH).

On peut parfois se servir du plan factoriel  $2 \times 2$  pour utiliser efficacement des sujets d'essai clinique en évaluant l'efficacité des deux traitements avec le même nombre de sujets que nécessiterait l'évaluation de l'efficacité d'un seul traitement. Cette application s'est avérée particulièrement utile pour des essais de taille importante relatifs à la mortalité. L'efficacité et la validité de cette application dépendent de l'absence d'interaction entre les traitements A et B, de sorte que les effets de A et de B sur les principales variables d'efficacité suivent un modèle additif et, donc, l'effet de A demeure virtuellement le même, qu'il s'ajoute ou non à l'effet de B. Comme pour l'essai croisé, on devra établir au préalable si ces conditions peuvent être respectées en se servant de données et d'informations *a priori*.

### 3.2 Essais multicentriques

On effectue des essais multicentriques pour deux grandes raisons. Premièrement, l'essai multicentrique est un moyen reconnu d'évaluer plus efficacement un nouveau médicament; dans certains cas, ce peut être le seul moyen de réunir un nombre suffisant de sujets pour réaliser l'objectif de l'essai dans un délai raisonnable. Les essais de ce genre peuvent, en principe, être effectués à n'importe quelle étape de la mise au point clinique. Il peut y avoir plusieurs centres comprenant un grand nombre de sujets, ou, s'il s'agit d'une maladie rare, un grand nombre de centres comprenant très peu de sujets.

Deuxièmement, l'essai multicentrique (ou essai multi-chercheurs) peut servir principalement à mieux asseoir la généralisation des résultats obtenus; cette application découle de la possibilité de recruter des sujets parmi une population plus vaste et d'administrer le médicament dans des

milieux cliniques plus variés, dans des conditions expérimentales plus proches de celles de l'utilisation future. En outre, la présence de plusieurs investigateurs peut avoir pour effet d'élargir l'éventail des évaluations cliniques relatives au médicament. Les essais de ce genre seraient des essais confirmatoires dans les dernières phases de la mise au point d'un médicament et ils impliqueraient vraisemblablement la participation d'un grand nombre de chercheurs et de centres. Ils pourraient parfois être menés dans divers pays afin de pousser plus loin la généralisation (voir le glossaire).

Pour pouvoir interpréter et extrapoler utilement les résultats d'un essai multicentrique, il faut que le mode d'application du protocole soit clair et qu'il soit identique partout. En outre, le calcul de la puissance et de la taille de l'échantillon repose sur l'hypothèse que les différences entre les traitements dans les centres sont des estimations non biaisées de la même quantité. Il est important de définir un protocole commun et de mener les essais en ayant à l'idée ces principes de base. On doit normaliser le plus entièrement possible les procédés. On peut uniformiser le plus possible les critères et les plans d'évaluation grâce à des réunions de chercheurs, à la formation préalable de membres du personnel et à un contrôle rigoureux durant l'essai. Un bon plan d'essai aura normalement pour but de répartir les sujets entre les traitements de la même façon dans chaque centre, et une bonne gestion aura pour but de soutenir cet objectif. Si on doit éventuellement tenir compte de l'hétérogénéité de l'effet du traitement d'un centre à l'autre, on aura avantage à opter pour des essais où le nombre des sujets ne varie pas excessivement d'un centre à l'autre et à éviter les essais qui impliquent la participation d'un petit nombre de très petits centres pour réduire l'écart entre les différentes estimations pondérées de l'effet du traitement. (Cette remarque ne concerne pas les essais qui impliquent uniquement la participation de très petits centres, ni ceux où le centre ne participe pas à l'analyse.) Si on négligeait de respecter ces conditions et si des doutes se formaient à propos de l'homogénéité des résultats, l'essai multicentrique pourrait, dans les cas extrêmes, être déprécié à un point tel qu'on le jugerait peu utile pour soutenir les revendications du promoteur.

Dans la version la plus élémentaire de l'essai multicentrique, chaque investigateur est responsable des sujets recrutés à un hôpital, de sorte que le «centre» est identifié uniquement au chercheur ou à l'hôpital. Pour de nombreux essais, toutefois, la situation est plus complexe. Par exemple, le chercheur peut recruter des sujets dans plusieurs hôpitaux, ou il peut représenter une équipe de cliniciens qui recrutent des sujets à leur propre clinique, à un hôpital ou à plusieurs hôpitaux affiliés. Si jamais la définition du centre dans un modèle statistique ne devait pas être claire, on se référera à la section statistique du protocole (voir section 5.1), qui doit définir clairement le terme (p. ex., selon le chercheur, l'endroit ou la région) pour l'essai concerné. Dans la plupart des cas, on peut définir convenablement le centre par rapport au chercheur, et la ligne directrice E6 de l'ICH contient des renseignements pertinents à cet égard. En cas de doute, on devra chercher à définir les centres de manière à garantir l'homogénéité

des principaux facteurs qui influent sur les mesures des variables primaires et de l'effet des traitements. Dans la mesure du possible, on devra indiquer et justifier dans le protocole toute disposition visant à grouper des centres pour l'analyse, mais quoi qu'il en soit, les décisions à ce sujet devront toujours être prises sans que l'on connaisse le traitement, par exemple au moment de l'examen aveugle.

Le protocole doit décrire le modèle statistique qui sera retenu pour estimer et tester les effets des traitements. On peut tout d'abord étudier l'effet majeur du traitement à l'aide d'un modèle qui prend en compte les différences des centres, mais qui ne contient pas de terme d'interaction traitement x centre. Si l'effet du traitement est homogène d'un centre à l'autre, l'inclusion systématique de termes d'interaction dans le modèle réduit l'efficacité du test pour les effets majeurs. Par ailleurs, si l'effet est réellement hétérogène, l'interprétation de l'effet majeur prête à la controverse.

Dans certains essais, par exemple certains essais de taille importante relatifs à la mortalité où il y a très peu de sujets dans chaque centre, il n'y aurait aucune raison de penser que les centres puissent avoir une influence quelconque sur les variables primaires ou secondaires, parce qu'il est peu probable qu'ils aient des effets importants du point de vue clinique. Dans d'autres essais, on pourrait reconnaître d'emblée que le nombre limité de sujets dans chaque centre empêche de tenir compte des effets du centre dans le modèle statistique. Dans ce cas, il sera inutile d'inclure dans le modèle un terme pour le centre et il ne sera pas nécessaire de stratifier la randomisation selon le centre.

Si on observe des effets favorables dans un essai où le nombre de sujets est élevé dans chaque centre, on devra normalement examiner si les effets des traitements dans les différents centres sont hétérogènes, car l'hétérogénéité peut nuire à la généralisation des conclusions de l'essai. On peut déceler l'existence d'une forte hétérogénéité à l'aide de graphiques reproduisant les résultats de chaque centre ou à l'aide de méthodes analytiques, comme l'application d'un test de signification à l'interaction traitement x centre. Lorsqu'on utilise un test de ce genre, il est important de reconnaître qu'il a généralement peu de puissance dans un essai conçu pour déceler l'effet majeur du traitement.

Si on constate que les effets des traitements sont hétérogènes, on devra user de circonspection et tenter de voir si cette hétérogénéité ne s'expliquerait pas par certains aspects de la gestion de l'essai ou par des caractéristiques des sujets. Si tel est le cas, il sera indiqué de pousser plus loin l'analyse et l'interprétation. Si, par contre, on ne peut trouver d'explication à cette hétérogénéité, illustrée par exemple par de fortes interactions quantitatives (voir le glossaire), il faudra peut-être estimer autrement l'effet du traitement, par une pondération des centres, dans le but de démontrer la robustesse de ces estimations. Il est encore plus important de

comprendre l'origine de l'hétérogénéité caractérisée par de fortes interactions qualitatives (voir le glossaire), et le fait de ne pas pouvoir expliquer cette hétérogénéité peut obliger l'observateur à mener d'autres essais cliniques avant de pouvoir estimer avec précision l'effet du traitement.

Jusqu'à maintenant, l'étude des essais multicentriques a reposé uniquement sur l'utilisation de modèles à effets fixes. On peut aussi se servir de modèles mixtes pour étudier l'hétérogénéité des effets des traitements. Ces modèles définissent les effets du centre et ceux de l'interaction traitement x centre comme aléatoires, et ils sont particulièrement utiles lorsque le nombre d'établissements est élevé.

### **3.3 Type de comparaison**

#### ***3.3.1 Essais visant à montrer une relation de supériorité***

Du point de vue scientifique, la façon la plus convaincante de montrer l'efficacité d'un médicament est de démontrer sa supériorité par rapport à un placebo, dans un essai contrôlé contre placebo, ou par rapport à une substance active-témoin, ou de démontrer une relation dose-effet. On appelle ce type d'essai «test de supériorité» (voir le glossaire). Dans la présente ligne directrice, il sera toujours question de test de supériorité, à moins d'indication contraire.

Lorsqu'il existe, pour des maladies graves, un traitement thérapeutique dont l'efficacité a été démontrée par un ou plusieurs tests de supériorité, on peut juger contraire à l'éthique la conduite d'un essai contrôlé contre placebo. On doit alors envisager l'utilisation rigoureusement scientifique d'une substance active-témoin. On décidera de la pertinence de mener un essai contrôlé contre placebo ou un essai contrôlé contre substance active selon les circonstances.

#### ***3.3.2. Essais visant à montrer une relation d'équivalence ou de non-infériorité***

Il arrive que l'on compare le produit à l'étude avec le traitement de référence sans avoir pour objectif d'en démontrer la supériorité. Ce type d'essai peut être assimilé à l'une ou l'autre de deux grandes catégories, selon l'objectif qu'on poursuit: essai d'«équivalence» (voir le glossaire) et test de «non-infériorité» (voir le glossaire).

Les essais de bioéquivalence appartiennent à la première catégorie. On effectue parfois des essais d'équivalence clinique pour d'autres motifs ayant trait à la réglementation

(par exemple, démontrer la relation d'équivalence clinique entre un médicament générique et le médicament commercialisé lorsque le composé n'est pas absorbé par l'organisme et qu'il n'est donc pas présent dans la circulation sanguine).

De nombreux essais contrôlés contre substance active ont pour but de montrer que le produit à l'étude n'est pas moins efficace que le traitement de comparaison, et ils font donc partie de la seconde catégorie. Il y a aussi les essais qui servent à comparer plusieurs doses du médicament à l'étude avec la dose recommandée, ou plusieurs doses, du médicament courant. Ces essais visent simultanément à montrer une relation dose-effet pour le médicament à l'étude et à comparer ce produit à la substance active-témoin.

Les essais contrôlés contre substance active, qu'ils appartiennent à l'une ou l'autre des deux catégories, peuvent aussi inclure un placebo, de sorte que l'on peut poursuivre plusieurs objectifs avec le même essai; par exemple, ils peuvent servir à établir la supériorité du médicament à l'étude par rapport au placebo et, donc, à valider le plan de l'essai et, en même temps, à comparer l'efficacité et l'innocuité du médicament à l'étude avec celles du comparateur actif. Les problèmes liés à la conduite des essais contrôlés contre substance active (essais d'équivalence ou tests de non-infériorité) qui n'incluent pas un placebo ou qui ne prévoient pas l'utilisation de doses multiples du nouveau médicament sont bien connus. Ces problèmes tiennent à l'absence implicite de mesures de validation interne (par opposition aux tests de supériorité), ce qui rend nécessaire la validation externe. L'essai d'équivalence (ou le test de non-infériorité) n'est pas conservatif par définition, de sorte que de nombreuses lacunes dans le plan et la conduite de l'essai tendront à biaiser les résultats dans le sens d'une conclusion d'équivalence. C'est pourquoi on doit accorder une attention toute spéciale au plan et à la conduite de ces essais. Ainsi, il est particulièrement important de réduire au minimum l'incidence du non-respect des critères d'admissibilité, de l'inobservation des directives, des retraits, des suivis irréalisables, des données manquantes et des autres écarts au protocole et de limiter le plus possible l'effet de ces facteurs sur les analyses pertinentes.

On doit choisir avec soin les comparateurs actifs. Par exemple, on pourrait choisir comme comparateur actif convenable une thérapie largement répandue dont l'efficacité dans l'indication considérée a été clairement établie et mesurée par un ou plusieurs tests de supériorité bien conçus et bien documentés, et qui devrait normalement être tout aussi efficace dans l'essai envisagé. C'est pourquoi on devra retrouver dans le nouvel essai les mêmes éléments de fond (variables primaires, dose du comparateur actif, critères d'admissibilité, etc.) qui ont caractérisé les tests de supériorité qui ont servi à

démontrer clairement l'efficacité clinique du comparateur actif, compte tenu de l'évolution des pratiques médicales et des méthodes statistiques ayant rapport au nouvel essai.

Le protocole d'un essai visant à démontrer une relation d'équivalence ou de non-infériorité doit absolument stipuler que c'est là le but explicite de l'essai. On doit aussi préciser un intervalle d'équivalence, qui est défini comme l'écart le plus grand qui puisse être jugé cliniquement acceptable. Cet intervalle doit être inférieur aux écarts observés dans les tests de supériorité appliqués au comparateur actif. En ce qui concerne les essais d'équivalence contre substance active, on doit se servir des bornes supérieure et inférieure de l'intervalle d'équivalence, tandis que la borne inférieure suffit pour les tests de non-infériorité contre substance active. On doit justifier cliniquement le choix de l'intervalle d'équivalence.

L'analyse statistique repose habituellement sur l'utilisation d'intervalles de confiance (voir section 5.5). On doit se servir d'intervalles bilatéraux pour les essais d'équivalence. On conclut à l'équivalence si l'intervalle de confiance est intégralement inclus dans l'intervalle d'équivalence. En pratique, cela revient à appliquer deux tests unilatéraux simultanés pour tester l'hypothèse nulle (composite) que la différence des traitements déborde l'intervalle d'équivalence contre l'hypothèse alternative (composite) que la différence des traitements est incluse intégralement dans l'intervalle. Comme les deux hypothèses nulles sont disjointes, on limite convenablement l'erreur de première espèce. En ce qui concerne les tests de non-infériorité, l'intervalle unilatéral s'impose. On utilise alors un test unilatéral pour tester l'hypothèse nulle que la différence des traitements (produit à l'étude moins témoin) est égale à la borne inférieure de l'intervalle d'équivalence contre l'hypothèse alternative que la différence des traitements est plus grande que la borne inférieure de l'intervalle d'équivalence. Le choix de la valeur de l'erreur de première espèce ne doit pas dépendre du genre de test utilisé: unilatéral ou bilatéral. Le calcul de la taille de l'échantillon doit reposer sur ces méthodes (voir section 3.5). Par ailleurs, il n'est pas opportun de conclure à l'équivalence ou à la non-infériorité sur la foi d'un résultat non significatif du test appliqué à l'hypothèse nulle qu'il n'y a pas de différence entre le produit à l'étude et le comparateur actif.

Le choix des ensembles d'analyse soulève aussi des questions particulières. Par exemple, si des sujets quittent le groupe de traitement ou le groupe de comparaison avant la fin de l'essai, le nombre des réponses sera vraisemblablement moins élevé et, donc, l'utilisation de l'ensemble d'analyse intégral (voir le glossaire) tendra à biaiser les résultats dans le sens d'une conclusion d'équivalence (voir section 5.2.3).

### **3.3.3 Essais visant à montrer une relation dose-effet**

On peut déterminer la nature de la relation dose-effet pour un nouveau produit de recherche à toutes les étapes de la mise au point et selon diverses méthodes (voir la ligne directrice E4 de l'ICH). Les essais visant à montrer une relation dose-effet poursuivent plusieurs objectifs, dont certains ont plus d'importance que d'autres: confirmation de l'efficacité, étude de la forme et de l'emplacement de la courbe dose-effet, estimation de la dose de départ appropriée, définition des stratégies optimales pour l'ajustement des doses individuelles, détermination de la dose au delà de laquelle le médicament a peu de chances de procurer un bienfait additionnel (dose maximale). On tentera de réaliser ces objectifs à l'aide des données recueillies pour un certain nombre de doses étudiées, y compris un placebo (dose zéro) si les circonstances le permettent. Dans ces conditions, l'application de méthodes visant à estimer la relation dose-effet, y compris la construction d'intervalles de confiance et l'utilisation de méthodes graphiques, est aussi importante que l'utilisation de tests statistiques. On pourrait devoir adapter les tests d'hypothèses utilisés à l'ordre naturel des doses ou à des questions particulières concernant la forme de la courbe dose-effet (ex.: monotonie). Le protocole doit décrire en détail les méthodes statistiques que l'on prévoit utiliser.

## **3.4 Essais séquentiels par groupes**

Les essais séquentiels par groupes servent à faciliter la réalisation des analyses intermédiaires (voir la section 4.5 et le glossaire). Bien que ce type d'essai ne soit pas le seul qui convienne pour les analyses intermédiaires, il est celui le plus couramment utilisé, car il est plus commode d'évaluer l'effet d'un médicament chez un groupe de sujets à intervalles réguliers pendant l'essai que de l'évaluer au fur et à mesure que des données sont produites sur chaque sujet. On doit définir entièrement les méthodes statistiques avant de connaître quoi que ce soit au sujet des effets du traitement et des affectations de traitements (ex. : décodage de l'insu, voir section 4.5). On peut confier au Comité indépendant de surveillance des données (voir le glossaire) le soin de faire une analyse intermédiaire des données d'un essai séquentiel par groupes ou le soin d'évaluer l'analyse faite par d'autres (voir section 4.6). Ce type de plan est surtout employé, avec succès d'ailleurs, dans les essais de longue durée et de taille importante

ayant trait à la mortalité ou aux principaux paramètres mesurés non mortels, mais il est de plus en plus utilisé à d'autres fins. En particulier, compte tenu de la nécessité reconnue de surveiller l'innocuité dans tous les essais, on doit toujours prévoir des modalités d'interruption de traitement pour la sécurité des sujets.

### 3.5 Taille de l'échantillon

On ne pourra répondre correctement à la question étudiée que si le nombre de sujets nécessaire pour un essai clinique est suffisamment élevé. Ce nombre est déterminé normalement par l'objectif primaire de l'essai; s'il l'est par un quelconque autre facteur, on doit le préciser clairement et le justifier. Par exemple, si l'effectif d'un échantillon de sujets est déterminé selon des normes d'innocuité ou des objectifs secondaires majeurs, il sera probablement plus élevé que l'effectif d'un échantillon fondé sur l'efficacité primaire (voir, par exemple, la ligne directrice E1A de l'ICH).

Si l'on utilise la méthode de calcul habituelle, on doit définir les éléments suivants: une variable primaire, la variable à tester, l'hypothèse nulle, l'hypothèse alternative (ou «de travail») pour la dose ou les doses choisies (y compris la différence de traitements dont on peut déceler l'existence, ou à l'inverse ne pas la reconnaître, pour la dose et la population de sujets désignées), la probabilité de rejeter par erreur l'hypothèse nulle (erreur de première espèce), et la probabilité de ne pas rejeter l'hypothèse nulle, alors qu'il aurait fallu (erreur de seconde espèce), ainsi que la manière de composer avec les retraits de sujets et les écarts au protocole. Dans certains cas, le taux d'effets est de toute première importance pour le calcul de la puissance, et on devrait formuler des hypothèses qui permettent de déterminer la taille de l'échantillon pour un essai à partir du nombre d'événements requis.

Le protocole doit décrire la méthode de calcul de la taille de l'échantillon et donner, s'il y a lieu, la valeur estimée des paramètres qui entrent dans le calcul (p. ex., variance, moyenne, taux de réponse, taux d'effets, différence dont il faut déceler l'existence). On doit aussi expliquer l'origine de ces estimations. Il est important de déterminer comment des dérogations aux hypothèses peuvent influencer sur l'estimation de la taille de l'échantillon; on peut faciliter cette analyse en considérant un éventail de tailles d'échantillon par rapport à un éventail raisonnable de dérogations aux hypothèses. Dans les essais confirmatoires, les hypothèses doivent normalement reposer sur des données publiées ou les résultats d'essais antérieurs. L'existence d'une différence de traitements pourra être décelée à partir d'un jugement concernant l'effet minimal qui a une pertinence clinique dans la gestion des patients ou un jugement concernant l'effet escompté du nouveau traitement, si celui-là est plus marqué. Habituellement, on fixe la probabilité de l'erreur de première espèce à 5 % ou moins, ou à toute autre valeur que pourraient dicter des ajustements rendus nécessaires par des considérations relatives à la



multiplicité; le choix pourra être influencé par la plausibilité *a priori* de l'hypothèse à tester et l'effet souhaité des résultats. On fixe ordinairement à un pourcentage allant de 10 à 20 % la probabilité de l'erreur de seconde espèce; le promoteur a intérêt à maintenir ce pourcentage le plus bas possible, surtout dans le cas des essais qu'il est difficile, sinon impossible, de répéter. On pourra, par ailleurs, choisir d'autres valeurs que les valeurs classiques de la probabilité de l'erreur de première espèce ou de seconde espèce; il sera même parfois souhaitable de le faire.

Le calcul de la taille de l'échantillon renvoie au nombre de sujets nécessaire pour l'analyse principale. S'il est question de l'«ensemble d'analyse intégral», on pourrait devoir rabaisser la taille d'effet estimée par comparaison à l'ensemble conforme au protocole (voir le glossaire), afin de tenir compte de la «dilution» de l'effet du traitement causée par la prise en compte de données concernant les patients qui se retirent de l'essai ou qui font preuve de peu de collaboration. Par ailleurs, on devra peut-être revoir les hypothèses relatives à la variabilité.

Dans un essai d'équivalence ou un test de non-infériorité (voir section 3.3.2), on devra normalement déterminer la taille de l'échantillon en ayant pour objectif de construire un intervalle de confiance pour la différence de traitements qui montre tout au plus un écart cliniquement acceptable entre les traitements. Lorsqu'on détermine la puissance d'un essai d'équivalence avec une différence vraie de zéro, la taille d'échantillon nécessaire pour réaliser cette puissance sera sous-estimée si la différence vraie n'est pas zéro. Par ailleurs, lorsqu'on détermine la puissance d'un test de non-infériorité avec une différence nulle, la taille d'échantillon nécessaire pour réaliser cette puissance sera sous-estimée si l'effet du produit à l'étude est moindre que celui de la substance active-témoin. On devra justifier la définition d'un écart «cliniquement acceptable» au regard de sa signification pour les patients futurs, et cet écart pourra être moindre que l'écart «cliniquement significatif» qui a été évoqué plus haut par rapport aux tests de supériorité qui visent à montrer l'existence d'une différence.

Dans les essais séquentiels par groupes, on ne peut établir à l'avance la taille exacte de l'échantillon parce que celle-ci dépend à la fois de l'effet du hasard, des règles d'interruption du traitement en vigueur et de la vraie différence de traitements. L'élaboration des règles d'interruption du traitement doit tenir compte de la répartition de l'échantillon, qui est exprimée habituellement dans la taille prévue et la taille maximum.

Lorsque le taux d'effets est moins élevé que prévu ou que la variabilité est plus grande que prévu, on peut utiliser méthodes pour recalculer la taille de l'échantillon sans dévoiler de données ni faire des comparaisons de traitements (voir section 4.4).

### **3.6 Saisie et traitement des données**

La collecte des données et leur transfert du chercheur au promoteur peuvent se faire par divers moyens: cahier d'observations, système de surveillance à distance, système informatique médical, transfert électronique. Quel que soit l'instrument utilisé pour la saisie des données, le fond et la forme des données recueillies doivent être en parfaite conformité avec le protocole et ils doivent être définis avant la conduite de l'essai clinique. On doit par ailleurs mettre l'accent sur les données nécessaires pour la mise en train de l'analyse prévue, y compris l'information de contexte (p. ex., fréquence des évaluations par rapport à la posologie) nécessaire pour confirmer le respect du protocole ou déceler les écarts majeurs au protocole. On doit pouvoir distinguer les «valeurs manquantes» de la «valeur nulle» et de la «caractéristique absente».

Toutes les étapes qui s'échelonnent de la saisie des données à la mise au point définitive de la base de données doivent être exécutées suivant la directive sur les bonnes pratiques cliniques (voir la ligne directrice E6 de l'ICH, section 5). En particulier, l'enregistrement des données et la correction des erreurs et des omissions doivent se faire par des procédés adéquats et fiables si l'on veut constituer une base de données de qualité et atteindre les objectifs de l'essai par la mise en train de l'analyse prévue.

## **4. CONSIDÉRATIONS RELATIVES À LA CONDUITE DES ESSAIS**

### **4.1 Surveillance des essais et analyse intermédiaire**

Le fait de mener un essai clinique en respectant soigneusement le protocole a une incidence majeure sur la crédibilité des résultats (voir la ligne directrice E6 de l'ICH). De même, une surveillance rigoureuse fera en sorte que l'on repérera très tôt les difficultés et on les éliminera dans la mesure du possible.

Deux types de surveillance caractérisent habituellement les essais cliniques confirmatoires menés par l'industrie pharmaceutique. Le premier a trait au contrôle de la qualité de l'essai, tandis que le second concerne le décodage de l'insu en vue de la comparaison de traitements (c.-à-d., analyse intermédiaire). Les deux types de surveillance diffèrent sur plus d'un plan: responsabilités différentes pour le personnel, accès à des types de renseignements (données d'essai et information) différents, et application de principes différents pour la limitation du biais statistique et opérationnel.

En ce qui regarde le contrôle de la qualité de l'essai, la surveillance consiste notamment à vérifier la pertinence des hypothèses du plan de l'essai, à vérifier si le protocole est respecté et si les données recueillies sont acceptables, à évaluer dans quelle mesure les objectifs de collecte

des données sont atteints et dans quelle mesure on réussit à maintenir l'effectif de l'échantillon soumis à l'essai (voir sections 4.2 à 4.4), etc. Ce type de surveillance n'implique aucunement l'accès aux données sur les effets comparés des traitements ni le dévoilement de données, et il n'a donc aucun effet sur l'erreur de première espèce. Il est de la responsabilité du promoteur (voir la ligne directrice E6 de l'ICH) et il peut être accompli par le promoteur lui-même ou un groupe indépendant désigné par ce dernier. En règle générale, ce type de surveillance débute au moment du choix des lieux de l'essai et se termine au moment de la collecte et de l'épuration des données relatives au dernier sujet.

L'autre type de surveillance (analyse intermédiaire) concerne le rassemblement des résultats comparatifs des traitements. L'analyse intermédiaire exige le décodage des données relatives à l'affectation des groupes de traitement (affectation des traitements ou identification des groupes de traitement) et des données récapitulatives comparatives sur les groupes de traitement. Le protocole doit donc contenir - sous peine d'être modifié au préalable - un plan d'analyse statistique destiné à l'analyse intermédiaire afin d'éviter certains types de biais. C'est ce dont il est question dans les sections 4.5 et 4.6.

#### **4.2 Modification des critères d'inclusion et d'exclusion**

Comme le stipule le protocole, on doit conserver les mêmes critères d'inclusion et d'exclusion durant toute la période de recrutement des sujets. On peut parfois modifier des critères, par exemple dans des essais de longue durée, lorsque le suggèrent les nouvelles connaissances médicales issues des recherches extérieures ou des analyses intermédiaires. Des changements sont aussi possibles lorsque les préposés à la surveillance se rendent compte que l'on passe outre régulièrement aux critères d'admissibilité ou que le taux de recrutement est extrêmement faible à cause de critères trop restrictifs. On doit pouvoir modifier des critères sans décoder l'insu et on doit toujours décrire ces modifications par une modification au protocole qui rende compte des conséquences statistiques - comme les corrections de la taille de l'échantillon engendrées par des différences de taux d'effets - ou des modifications apportées à l'analyse prévue initialement, par exemple stratification de l'analyse selon les nouveaux critères d'inclusion ou d'exclusion.

#### **4.3 Taux de constitution de l'échantillon**

On doit suivre l'évolution du taux de constitution de l'échantillon dans les essais où la constitution de l'échantillon s'étend sur une longue période. Si ce taux chute bien au-dessous du niveau prévu, on doit pouvoir l'expliquer et adopter des mesures correctives afin de conserver

la puissance de l'essai et de dissiper les craintes concernant la non-objectivité des critères d'admissibilité et d'autres aspects de la qualité d'un essai. Dans les essais multicentriques, ces considérations incombent à chacun des centres d'essai.

#### **4.4 Correction de la taille de l'échantillon**

Dans les essais de longue durée, il nous est donné normalement de vérifier les hypothèses qui sous-tendent le plan initial et le calcul de la taille de l'échantillon. Cette opération peut être particulièrement importante si les caractéristiques de l'essai ont été définies en fonction de données provisoires ou incertaines. Si, par un contrôle provisoire appliqué aux données codées, on observe que la variance globale des réactions, le taux d'effets ou le taux de survie s'écartent des prévisions, on peut recalculer la taille de l'échantillon à l'aide de nouvelles hypothèses, en justifiant cette opération dans une modification au protocole et dans le rapport de l'étude clinique. On doit exposer les mesures prises pour maintenir l'insu et, s'il y a lieu, les conséquences de ces mesures pour l'erreur de première espèce et l'étendue des intervalles de confiance. Le protocole doit prévoir dans la mesure du possible la nécessité éventuelle de recalculer la taille de l'échantillon (voir section 3.5).

#### **4.5 Analyse intermédiaire et interruption du traitement**

Une analyse intermédiaire est une analyse qui a pour but de comparer des groupes d'essai au point de vue de l'efficacité ou de l'innocuité avant que l'essai ne prenne fin. Comme ces comparaisons influent sur l'interprétation des résultats de l'essai par leur nombre, leurs conséquences et les méthodes utilisées, on doit planifier soigneusement à l'avance toutes les analyses intermédiaires et les décrire dans le protocole. Dans des circonstances exceptionnelles, on pourra devoir exécuter une analyse intermédiaire qui n'était pas prévue à l'origine. On devra en l'occurrence rédiger une modification au protocole qui décrit cette analyse, avant que ne soient dévoilées les données comparatives des traitements. Lorsqu'on envisage de faire une analyse intermédiaire dans le but de décider si on met fin ou non à un essai, on utilise habituellement un plan séquentiel par groupes fondé sur des méthodes de contrôle statistique (voir section 3.4). Ce genre d'analyse intermédiaire a pour objectif de mettre fin prématurément à l'essai si la supériorité du traitement à l'étude a été clairement établie, s'il est devenu peu probable d'observer une différence de traitements significative ou si on constate des effets indésirables inacceptables. En règle générale, il faudra plus d'information pour mettre fin prématurément à un essai dans le cas du contrôle de l'efficacité que pour le contrôle de l'innocuité (c'est-à-dire que le contrôle de l'efficacité est plus restrictif). Lorsque le plan de l'essai et l'objectif de la surveillance font appel à plusieurs paramètres mesurés, on doit aussi tenir compte de cet aspect de la multiplicité.

Le protocole doit contenir le calendrier des analyses intermédiaires ou, à tout le moins, exposer les considérations qui guideront l'élaboration de ce calendrier (par exemple, si on se sert de méthodes fondées sur la fonction de dépense alpha, reconnues pour leur souplesse); d'autres détails pourront être inclus dans une modification au protocole avant que ne débute la première analyse intermédiaire. Le protocole, ou les modifications au protocole, devront exposer clairement les règles d'interruption de traitement et leurs caractéristiques. On devra aussi prendre en considération l'incidence probable de l'interruption du traitement sur l'analyse d'autres variables importantes. Ces remarques devront être rédigées ou approuvées par le Comité de surveillance des données (voir section 4.6), s'il en existe un pour cet essai. Si l'on s'écarte de la procédure prévue, on risque toujours d'invalider les résultats de l'essai. S'il devient nécessaire d'apporter des modifications à l'essai, à la première occasion on devra indiquer dans une modification au protocole les conséquences de ce changement pour les méthodes statistiques, en insistant sur l'effet probable de ces modifications sur les analyses et les inférences qui seront faites. Quelles que soient les méthodes retenues, on doit toujours veiller à limiter la probabilité globale de l'erreur de première espèce.

Les analyses intermédiaires doivent être faites sous le sceau de la plus stricte confidentialité, à cause du risque de décodage des données ou des résultats de l'essai. On ne doit pas dévoiler les résultats de ces analyses aux personnes chargées de la conduite de l'essai, car cela pourrait modifier leur attitude à l'égard de l'essai et les amener à réviser les caractéristiques recherchées chez un sujet ou à fausser les comparaisons de traitements. Cette règle vise tous les investigateurs ainsi que les employés du promoteur, à l'exception de ceux qui ont la responsabilité directe des analyses intermédiaires. Seules les décisions concernant l'interruption d'un essai ou la modification des modalités de l'essai doivent être communiquées aux investigateurs.

Dans la plupart des essais cliniques qui visent à confirmer l'efficacité et l'innocuité d'un produit de recherche, on doit chercher à constituer entièrement l'échantillon prévu à l'origine; on devrait mettre fin prématurément à un essai uniquement pour des raisons d'éthique ou si la puissance de l'essai a atteint un niveau inacceptable. Par ailleurs, il est admis que dans les projets de mise au point de médicaments, le promoteur doit pouvoir avoir accès aux données comparatives des traitements pour diverses raisons (p. ex., planification d'autres essais). Il est également admis que seul un sous-ensemble des essais servira à étudier des maladies graves ou la mortalité et qu'il faudra probablement, pour des raisons d'éthique, contrôler par étapes le rassemblement des résultats comparatifs des traitements. Dans l'un et l'autre cas, on devra avoir défini dans le protocole, ou dans des modifications au protocole, des plans d'analyse statistique avant que ne soient décodées les données comparatives des traitements, dans le but de réduire les risques de biais statistique ou opérationnel.

Pour de nombreux essais cliniques portant sur des produits de recherche, en particulier les essais qui ont une grande importance du point de vue de la santé publique, on devra confier la surveillance des comparaisons des données sur l'efficacité et l'innocuité à un groupe de l'extérieur désigné souvent comme le comité indépendant de surveillance des données (CISD) (ou comité de surveillance de l'innocuité et des données, ou encore comité de surveillance des données), dont les attributions devront être clairement définies.

Lorsqu'un promoteur a pour rôle de surveiller la comparaison des données sur l'efficacité et l'innocuité et qu'il a accès par conséquent à des données comparatives décodées, on doit veiller à préserver l'intégrité de l'essai et à gérer prudemment la diffusion de l'information. Le promoteur doit voir à ce que le comité de surveillance interne respecte les modes opératoires normalisés et à ce que soit rédigé le procès-verbal des réunions de prise de décision, y compris des rapports contenant les résultats provisoires, et il doit en rendre compte officiellement.

Une analyse intermédiaire mal planifiée (qui aurait ou non pour conséquence d'interrompre prématurément l'essai) peut fausser les résultats de l'essai et affecter vraisemblablement la validité des conclusions. Il faut donc éviter de faire de telles analyses, mais dans le cas contraire on devra expliquer dans le rapport d'étude clinique pourquoi on a dû faire une analyse non planifiée et indiquer dans quelle mesure l'insu a été levé; on devra aussi évaluer la grandeur probable du biais qui a été introduit, ainsi que l'effet sur l'interprétation des résultats.

#### **4.6 Rôle du Comité indépendant de surveillance des données (CISD) (voir sections 1.25 et 5.5.2 de la ligne directrice E6 de l'ICH)**

Le promoteur peut mettre sur pied un CISD qui aura pour mandat d'évaluer à intervalles réguliers la progression de l'essai clinique, les données sur l'innocuité, ainsi que les principales variables d'efficacité, et d'indiquer au promoteur s'il doit poursuivre l'essai, lui apporter des modifications ou y mettre fin. Le Comité devra avoir des règles de fonctionnement et rédiger un procès-verbal à chacune de ses réunions. Il devra en outre consigner les résultats provisoires de l'essai. Ces données devront être accessibles pour examen une fois l'essai terminé. L'indépendance du Comité vise à garantir le contrôle de la diffusion de données comparatives importantes et à préserver l'intégrité de l'essai clinique, qui pourrait souffrir des conséquences négatives de l'accès aux données de l'essai. Le CISD n'est pas un comité d'examen de l'établissement (CEÉ), ni un comité d'éthique indépendant (CÉI) et il compte normalement parmi ses membres des spécialistes des essais cliniques versés dans les disciplines pertinentes, y compris la statistique.

Si des représentants du promoteur siègent au CISD, leur rôle doit être clairement défini dans les règles de fonctionnement du comité (par exemple, droit de vote sur les questions clés?). Comme ces représentants peuvent avoir accès à des données décodées, on doit prévoir des dispositions pour le contrôle de la diffusion des résultats provisoires des essais au sein de l'organisation du promoteur.

## 5. CONSIDÉRATIONS RELATIVES À L'ANALYSE DES DONNÉES

### 5.1 Définition préalable de l'analyse

Lorsqu'on élabore le plan d'un essai clinique, on doit inclure dans la section statistique du protocole les principales caractéristiques de l'analyse statistique prévue. Cette section doit énumérer toutes les caractéristiques majeures de l'analyse confirmatoire des variables primaires et indiquer de quelle manière seront résolus les problèmes d'analyse qui pourraient surgir. En ce qui concerne les essais exploratoires, cette section pourra exposer des orientations et des principes plus généraux.

Le plan d'analyse statistique (voir le glossaire) est un document qui peut être rédigé à part, et que l'on peut compléter après que le protocole a été mis au point. Le plan d'analyse statistique pourra contenir une description plus technique et plus détaillée des principaux points énoncés dans le protocole (voir section 7.1), ainsi que la description de la méthode d'analyse des variables primaires et secondaires et des autres données. On devra réexaminer et, peut-être, mettre à jour le plan par suite de l'examen aveugle des données (voir la section 7.1 pour la définition) et on devra l'achever avant le décodage de l'insu. Il faudra en outre consigner la date à laquelle le plan a été achevé et la date du décodage de l'insu.

Si l'examen aveugle indique qu'il faut modifier les principaux points du protocole, les changements pertinents doivent être décrits dans une modification au protocole. Autrement, il suffira de mettre à jour le plan d'analyse statistique en tenant compte des considérations découlant de l'examen aveugle. Seuls les résultats des analyses prévues dans le protocole (et dans les modifications au protocole) peuvent être considérés comme des résultats confirmatoires.

Enfin, la section statistique du rapport d'études cliniques doit contenir une description claire de la méthodologie statistique et indiquer à quel moment dans l'élaboration de l'essai clinique on a décidé de la méthodologie (voir la ligne directrice E3 de l'ICH).

## 5.2 Ensembles d'analyse

L'ensemble des sujets dont les données feront l'objet des principales analyses doit être défini dans la section statistique du protocole. En outre, il sera utile de disposer de renseignements de base sur tous les sujets inscrits à un essai (p. ex., période d'adaptation). Le type de renseignements recherchés dépendra des caractéristiques de l'essai, mais on s'efforcera de recueillir à tout le moins des données démographiques et des données de base sur l'état de la maladie.

Si tous les sujets randomisés dans un essai clinique satisfaisaient à tous les critères d'admissibilité, respectaient parfaitement toutes les étapes de l'essai (sans aucun suivi irréalisable) et produisaient un enregistrement complet, l'ensemble des sujets à analyser irait de soi. L'élaboration et la conduite d'un essai doivent avoir pour objectif ultime de réaliser ces conditions, mais cette éventualité est très peu probable dans la pratique. Par conséquent, la section statistique du protocole doit décrire l'effet que ce problème pourra avoir éventuellement sur les sujets et les données à analyser. Le protocole doit en outre définir une méthode pour réduire au minimum le nombre d'irrégularités dans la conduite de l'essai qui pourraient nuire à la qualité de l'analyse (p. ex., écarts au protocole, retraits, valeurs manquantes). Il doit proposer des façons de réduire la fréquence de ces difficultés et de résoudre les problèmes qui surgissent dans l'analyse des données. Durant l'examen aveugle on pourra envisager des modifications dans la manière dont l'analyse traite les écarts au protocole. Il est souhaitable de définir chaque écart au protocole d'importance majeure en fonction du moment où il est survenu, de sa cause et de son effet sur les résultats de l'essai. Le rapport de l'étude clinique doit mentionner la fréquence des écarts au protocole et le type d'écart, décrire les valeurs manquantes et les autres problèmes et exposer l'effet probable de ces facteurs sur les résultats de l'essai (voir la ligne directrice E3 de l'ICH).

Les décisions concernant l'ensemble d'analyse doivent être fondées sur les principes suivants:  
1) réduction maximale du biais, et 2) limitation de l'erreur de première espèce.

### 5.2.1 Ensemble d'analyse intégral

Le principe de l'intention de traiter (voir le glossaire) suppose que l'analyse primaire doit s'appliquer à tous les sujets randomisés. La stricte application de ce principe nécessiterait de faire le suivi intégral de tous les sujets randomisés pour les besoins de l'essai. Or, cela est difficilement réalisable dans la pratique pour des raisons qui restent à exposer. Dans ce document, le terme «ensemble d'analyse intégral» désigne l'ensemble d'analyse qui est le plus complet possible et qui se rapproche le plus du principe de l'intention de traiter. Préserver la randomisation initiale dans l'analyse est un



moyen précieux de limiter le biais et d'asseoir solidement les tests statistiques. Dans beaucoup d'essais cliniques, l'utilisation de l'ensemble d'analyse intégral est une stratégie prudente. Elle peut aussi, dans beaucoup de cas, produire des estimations des effets de traitement vraisemblablement plus proches de celles observées dans les essais ultérieurs.

On peut être amené à exclure des sujets randomisés de l'ensemble d'analyse intégral dans des cas particuliers, à savoir: le sujet ne satisfait pas aux principaux critères d'admissibilité; le sujet n'a pas pris une seule dose du médicament à l'essai; aucune post-randomisation des données. On doit toujours justifier ces cas d'exclusion. Seulement dans les circonstances suivantes peut-on exclure de l'analyse les sujets qui ne satisfont pas à tous les critères d'admissibilité sans risquer d'introduire un biais:

- 1) le critère d'admissibilité a été mesuré avant la randomisation;
- 2) on peut déterminer de façon parfaitement objective les critères qui ne sont pas respectés;
- 3) on prête la même attention à tous les sujets pour ce qui est de la détermination des critères non respectés; (Cette condition peut être difficile à respecter dans des essais ouverts, ou même dans des essais à double insu si les données sont décodées avant l'exercice, d'où l'importance de l'examen aveugle.)
- 4) tous les cas de non-respect des critères d'admissibilité qui ont été déterminés sont exclus.

Dans certains cas, il peut être raisonnable d'exclure de l'ensemble des sujets randomisés les sujets qui n'ont pas pris du tout le médicament à l'essai. Le principe de l'intention de traiter sera sauf malgré ces exclusions si, par exemple, la décision de débiter ou non le traitement n'a pas été influencée par la connaissance du traitement affecté. Dans d'autres cas, il faudra peut-être exclure de l'ensemble des sujets randomisés les personnes pour lesquelles il n'y a pas de post-randomisation des données. Pour qu'une analyse soit réputée complète, on devra avoir examiné au préalable la question des biais que peuvent engendrer ces exclusions ou d'autres types d'exclusions.

Lorsqu'on utilise l'ensemble d'analyse intégral, les écarts au protocole qui se manifestent après la randomisation peuvent avoir un effet sur les données et les conclusions de l'essai, surtout si leur manifestation a rapport à l'affectation des traitements. À bien des égards il convient néanmoins d'inclure dans l'analyse les données des sujets visés, en conformité avec le principe de l'intention de traiter. Des problèmes particuliers surgissent lorsque des sujets quittent après avoir reçu une ou plusieurs doses et qu'on ne peut plus recueillir de données sur eux par la suite, ou lorsqu'il n'est plus possible d'assurer le suivi de certains sujets, car la «non-inclusion» de ces sujets dans l'ensemble d'analyse intégral peut miner considérablement le processus. C'est pourquoi les mesures de variables primaires enregistrées au moment de l'interruption du suivi pour une raison quelconque, ou recueillies par la suite en conformité avec le calendrier prévu dans le protocole, sont précieuses dans ces circonstances; la seconde formule est particulièrement importante dans les études qui ont pour variable primaire la mortalité ou une maladie grave. Si on a l'intention de recueillir des données de cette manière, on doit le mentionner dans le protocole. Par ailleurs, on peut recourir à des méthodes d'imputation - depuis le report de la dernière observation jusqu'à l'utilisation de modèles mathématiques complexes - afin de compenser l'effet des données manquantes. Enfin, si on choisit de recourir à d'autres méthodes pour obtenir des mesures de variables primaires pour chaque membre de l'ensemble d'analyse intégral, on pourrait devoir poser des hypothèses concernant les effets des traitements ou une version plus simple de ces effets (p. ex., succès/échec). Si on décide d'employer l'une ou l'autre de ces méthodes dans l'analyse, on doit en faire la description, et en donner la justification, dans la section statistique du protocole et on doit exposer clairement les hypothèses qui sous-tendent les modèles mathématiques qui pourraient être utilisés. Il importe également de démontrer la robustesse des résultats d'analyse, surtout si la méthode proposée risque de produire des estimations biaisées des effets de traitement.

Étant donné l'imprévisibilité de certains problèmes, il est parfois souhaitable d'attendre le moment de l'examen aveugle des données, à la fin de l'essai, pour examiner en détail des pistes de solution possibles, et si tel est le cas, on doit le mentionner dans le protocole.

### 5.2.2 Ensemble conforme au protocole

L'ensemble de sujets «conforme au protocole», désigné parfois comme l'ensemble des «cas valides» ou l'échantillon «d'efficacité», ou encore l'échantillon des «sujets évaluables», est un sous-ensemble de l'ensemble d'analyse intégral qui comprend les sujets plus respectueux du protocole et qui est caractérisé par des critères comme les suivants:

- 1) le sujet doit avoir franchi un nombre minimum d'étapes selon le schéma thérapeutique, ce nombre ayant été défini à l'avance;
- 2) il doit exister des mesures des variables primaires;
- 3) on ne doit observer aucun écart majeur au protocole, et notamment aucun cas de non-respect des critères d'admissibilité.

Si on doit exclure des sujets de l'ensemble conforme au protocole, on devra en exposer les raisons exactes avant le décodage de l'insu d'une manière appropriée aux conditions de l'essai.

L'utilisation de l'ensemble conforme au protocole dans l'étude d'un nouveau traitement peut accroître les chances d'observer une plus grande efficacité pour ce traitement, et elle est aussi plus conforme au modèle scientifique sur lequel repose le protocole. Cependant, le test d'hypothèse correspondant peut ou non être conservatif, comme l'estimation de l'effet de traitement peut ou non être prudente, tout dépendant de l'essai; quant au biais, qui peut être notable, il découle du fait que le respect du protocole peut avoir un rapport avec le traitement et l'effet du traitement.

On doit définir clairement les problèmes qui forcent l'exclusion de sujets et, par conséquent, la constitution de l'ensemble conforme au protocole, ainsi que les autres écarts au protocole, parmi lesquels on retrouve les exemples suivants: erreur dans l'affectation des traitements, usage de médicaments non visés par l'étude, faible collaboration du sujet, suivi irréalisable, et données manquantes. Il est recommandé d'évaluer quelle est la fréquence de ces problèmes chez les groupes de traitement et à quel moment ils surviennent dans l'essai.

### 5.2.3 Fonctions des divers ensembles d'analyse

En règle générale, il est avantageux de démontrer que les principaux résultats de l'essai sont peu sensibles au choix de l'ensemble des sujets étudié. Dans les essais confirmatoires, il est normal de prévoir une analyse de l'ensemble intégral ainsi qu'une analyse de l'ensemble conforme au protocole, de sorte que l'on puisse examiner formellement les différences entre les deux analyses et en faire l'interprétation. Dans certains cas, il sera souhaitable d'examiner jusqu'à quel point les conclusions de l'essai peuvent être sensibles au choix de l'ensemble des sujets étudié. Lorsque l'ensemble intégral et l'ensemble conforme au protocole permettent de tirer essentiellement les mêmes conclusions, on reconnaît aux résultats de l'essai une plus grande validité, mais il faut se rappeler que si l'on devait exclure une forte proportion de sujets de l'ensemble conforme au protocole, la validité globale de l'essai serait remise en question.

L'ensemble d'analyse intégral et l'ensemble conforme au protocole ont des fonctions différentes dans les tests de supériorité (qui visent à démontrer la supériorité du produit à l'étude) et dans les essais d'équivalence ou les tests de non-infériorité (qui visent à démontrer l'équivalence du produit à l'étude, voir section 3.3.2). Dans les tests de supériorité, on utilise l'ensemble intégral pour l'analyse primaire (sauf dans des cas exceptionnels) parce qu'il ne mène pas à une surestimation de l'efficacité comme le fait l'ensemble conforme au protocole, étant donné que la présence, dans l'ensemble intégral, de sujets n'observant pas le protocole a généralement pour effet de réduire l'estimation de l'effet du traitement. Par contre, il n'est généralement pas prudent d'utiliser l'ensemble d'analyse intégral dans les essais d'équivalence ou les tests de non-infériorité, et on doit considérer très attentivement son rôle dans ces circonstances.

### 5.3 Valeurs manquantes et valeurs aberrantes

Les valeurs manquantes sont une source possible d'erreur systématique dans les essais cliniques. C'est pourquoi on doit faire tous les efforts nécessaires pour se conformer aux dispositions du protocole relatives à la collecte et à la gestion des données. Dans la réalité toutefois, on devra presque toujours composer avec l'absence partielle de données. Néanmoins, on pourra considérer comme valides des essais où des données seront manquantes, à la condition que les méthodes de traitement des valeurs manquantes soient rationnelles et, surtout, qu'elles soient décrites dans le protocole. Il y a moyen d'améliorer le processus de définition de ces méthodes en revoyant les sections pertinentes du plan d'analyse statistique au moment de l'examen aveugle. Malheureusement, on ne connaît pas de méthode

de traitement des valeurs manquantes d'application universelle. On devrait néanmoins examiner le degré d'influence du choix de la méthode de traitement des valeurs manquantes sur les résultats de l'analyse, surtout s'il y a beaucoup de valeurs manquantes.

On devra faire de même pour étudier l'influence des valeurs aberrantes, dont la définition statistique est jusqu'à un certain point arbitraire. On peut identifier clairement une valeur particulière comme valeur aberrante lorsqu'elle est définie comme telle du point de vue médical et du point de vue statistique; alors, le contexte médical indiquera souvent la méthode de traitement appropriée. On doit veiller à ce que les méthodes de traitement des valeurs aberrantes décrites dans le protocole ou le plan d'analyse statistique ne favorisent aucun groupe de traitement *a priori*. Là aussi, on peut revoir utilement cet aspect de l'analyse au moment de l'examen aveugle. Si le protocole de l'essai ne mentionne aucune méthode de traitement des valeurs aberrantes, on devra faire une analyse avec les valeurs réelles et au moins une autre, où on éliminera ou réduira l'effet des valeurs aberrantes, et on devra examiner les différences entre les résultats de ces deux analyses.

#### 5.4 Transformation de données

La décision de transformer des données relatives à des variables clés se prend idéalement au moment de la conception du plan de l'essai; on se fonde alors sur des données comparables tirées d'essais cliniques antérieurs. Le protocole doit mentionner le type de transformation choisi (p. ex., racine carrée, logarithme) et il doit justifier ce choix, surtout dans le cas des variables primaires. Les ouvrages de base exposent les principes généraux qui guident l'utilisation des transformations dans le respect des hypothèses sur lesquelles reposent les méthodes statistiques; par ailleurs, dans un certain nombre de champs d'études cliniques, des règles particulières ont été élaborées pour certaines variables. La décision de transformer les données relatives à une variable et le choix du type de transformation tiennent à la volonté d'utiliser une échelle qui facilite l'interprétation clinique.

Les mêmes remarques s'appliquent à l'utilisation d'autres variables calculées, comme la variation par rapport aux données de base, la variation en pourcentage par rapport aux données de base, la «surface sous la courbe» liée à des mesures répétées, ou le rapport de deux variables. On doit considérer soigneusement l'interprétation clinique correspondante et justifier les calculs dans le protocole. La section 2.2.2. traite des questions analogues.

## 5.5 Estimation, intervalles de confiance et tests d'hypothèses

La section statistique du protocole doit préciser les hypothèses à tester et les effets de traitement à estimer dans le but de répondre aux objectifs premiers de l'essai. On doit décrire les méthodes statistiques qui seront employées à cette fin pour les variables primaires (et, idéalement, les variables secondaires) et définir clairement le modèle statistique correspondant. On devra indiquer dans la mesure du possible un intervalle de confiance pour chaque estimation de l'effet de traitement et définir la méthode de calcul de ces intervalles. On devra aussi mentionner, s'il y a lieu, l'intention d'utiliser des données de base pour accroître la précision ou redresser des estimations en fonction des écarts probables, par exemple au moyen de l'analyse de covariance.

Il est important de préciser si on utilisera des tests de signification statistique unilatéraux ou bilatéraux et, en particulier, de justifier par anticipation l'utilisation de tests unilatéraux. Si les tests d'hypothèses ne sont pas jugés convenables, on devra décrire de quelle manière on entend arriver à des conclusions statistiques. La question des méthodes d'inférence à test unilatéral ou bilatéral est sujette à controverse et les ouvrages statistiques font état d'opinions diverses sur cette question. Dans les essais menés pour des motifs ayant trait à la réglementation, il est préférable de fixer la valeur de l'erreur de première espèce pour les tests unilatéraux à la moitié de la valeur classique de l'erreur de première espèce pour les tests bilatéraux, ce qui est davantage en rapport avec les intervalles de confiance bilatéraux, qui conviennent généralement pour estimer l'écart probable entre deux traitements.

Le modèle statistique retenu devra refléter l'état actuel des connaissances médicales et statistiques concernant les variables à analyser, ainsi que le plan de l'essai. On devra préciser tous les effets à ajuster dans l'analyse (par exemple, dans des modèles d'analyse de variance) et décrire, s'il y a lieu, de quelle manière on pourrait modifier cet ensemble d'effets après avoir reçu des résultats provisoires. Les mêmes remarques valent pour l'ensemble des covariables examinées dans une analyse de covariance (voir aussi section 5.7). Au moment du choix des méthodes statistiques, on doit être attentif à la distribution statistique des variables primaires et des variables secondaires. Par exemple, lorsqu'on doit choisir entre des méthodes paramétriques et des méthodes non paramétriques, il est important de se rappeler qu'il faut produire des estimations des effets de traitements, accompagnées des intervalles de confiance (outre les tests de signification).

On doit distinguer clairement l'analyse principale de la variable primaire et les analyses auxiliaires des variables primaires ou secondaires. De plus, la section statistique du protocole ou le plan d'analyse statistique devra décrire de quelle manière les données autres que celles

ayant trait aux variables primaires et secondaires seront résumées et présentées; on devra par la même occasion indiquer les moyens utilisés pour garantir la cohérence de l'analyse dans une série d'essais, par exemple en ce qui concerne les données sur l'innocuité.

Les modèles qui contiennent de l'information sur des paramètres pharmacologiques connus, le degré d'observation du protocole chez les sujets, ou d'autres données à caractère biologique peuvent nous renseigner grandement sur l'efficacité réelle ou potentielle d'un traitement, surtout en ce qui regarde l'estimation des effets de ce traitement. On doit toujours définir clairement les hypothèses qui sous-tendent ces modèles et décrire soigneusement les limites des conclusions qui peuvent en être tirées.

## **5.6 Modification du niveau de signification et du niveau de confiance**

Lorsqu'il y a multiplicité, la méthode d'analyse des données d'essais cliniques axée sur l'observation de fréquences peut nécessiter une modification de la valeur de l'erreur de première espèce. La multiplicité vient, par exemple, de l'existence de plusieurs variables primaires (voir section 2.2.2) ou de plusieurs comparaisons de traitements, d'évaluations répétées ou d'analyses intermédiaires (voir section 4.5). Il est parfois souhaitable de recourir à des méthodes qui fassent abstraction de la multiplicité ou qui l'amenuisent, lorsque c'est possible, par exemple identification de la variable primaire clé (variables multiples), choix d'une méthode critique de comparaison de traitements (comparaisons multiples), utilisation d'une mesure globale telle que la «surface sous la courbe» (mesures répétées). Dans le cas des essais confirmatoires, on doit mentionner dans le protocole tous les aspects de la multiplicité qui subsisteraient après que de telles méthodes ont été appliquées; par ailleurs, on doit toujours envisager la possibilité d'une modification, et on doit décrire dans le plan d'analyse toute modification éventuelle ou, à l'inverse, expliquer pourquoi on ne juge pas nécessaire d'opérer une modification.

## **5.7 Sous-groupes, interactions et covariables**

La ou les variables primaires ont souvent un rapport systématique avec d'autres facteurs que le traitement. Par exemple, elles peuvent être liées à des covariables comme l'âge et le sexe, ou il peut y avoir des différences entre des sous-groupes particuliers de sujets, comme ceux traités dans les différents centres qui participent à un essai multicentrique. Parfois, des modifications pour tenir compte de l'effet des covariables ou des effets des sous-groupes feront partie intégrante de l'analyse et devront par conséquent être indiquées dans le protocole. Dans les discussions pré-essai, on doit déterminer les covariables et les facteurs qui devraient avoir une influence notable sur la ou les variables primaires et on doit examiner comment tenir compte de ces covariables ou facteurs dans l'analyse de manière à accroître le degré de précision et à

compenser l'absence possible d'équilibre entre les groupes de traitement. Si on doit utiliser un ou plusieurs facteurs de stratification dans le plan de l'essai, il convient d'en tenir compte dans l'analyse. Lorsqu'on met en doute l'utilité d'une modification, il est souvent conseillé de porter toute son attention sur l'analyse sans modification, l'analyse avec modification ayant un rôle secondaire. On doit accorder une attention particulière aux effets du centre d'essai et au rôle des valeurs de base de la variable primaire. Il n'est pas recommandé de modifier les analyses principales en fonction de covariables mesurées après la randomisation, parce que ces valeurs peuvent être influencées par les traitements.

L'effet du traitement peut lui-même varier également selon le sous-groupe ou la covariable; par exemple, il peut diminuer à mesure que l'âge augmente ou il peut être plus élevé pour un groupe de sujets particulier défini selon le diagnostic. Parfois, on s'intéressera d'avance à ces interactions (p. ex., en gériatrie); en conséquence, l'analyse de l'essai confirmatoire comprendra une analyse de sous-groupe ou un modèle statistique avec interactions. Cependant, les analyses de sous-groupe ou les analyses d'interactions sont le plus souvent exploratoires et elles doivent être clairement identifiées comme telles; elles serviront surtout à étudier l'uniformité des effets du traitement observés dans l'ensemble. En règle générale, on exécutera ces analyses en incluant tout d'abord des termes d'interaction dans le modèle statistique utilisé, puis en appliquant d'autres analyses exploratoires à des sous-groupes de sujets pertinents ou à des strates définies par les covariables. On doit interpréter les analyses exploratoires avec circonspection; si on concluait à l'efficacité (ou à l'inefficacité) ou à l'innocuité d'un traitement en se fondant uniquement sur les résultats d'analyses exploratoires de sous-groupes, ces conclusions auraient peu de chances d'être acceptées.

## **5.8 Intégrité des données et validité des logiciels**

La crédibilité des résultats d'une analyse dépend de la qualité et de la validité des méthodes et des logiciels («maison» ou commerciaux) utilisés pour la gestion des données (saisie, stockage, vérification, correction et extraction) et leur traitement statistique. C'est pourquoi les opérations de gestion des données doivent être effectuées selon des modes opératoires normalisés complets et efficaces. On doit pouvoir compter sur des logiciels de gestion et d'analyse statistique fiables et avoir accès aux résultats des tests de fiabilité.



## 6. ÉVALUATION DE L'INNOCUITÉ ET DE LA TOLÉRABILITÉ

### 6.1 Portée de l'évaluation

L'évaluation de l'innocuité et de la tolérabilité (voir le glossaire) est une dimension importante de l'essai clinique. Dans les premières phases de l'essai, elle est surtout de nature exploratoire et porte uniquement sur la toxicité du médicament, alors que dans les phases ultérieures elle est plutôt caractérisée par l'étude d'échantillons plus vastes. Les essais contrôlés menés à des phases ultérieures sont une façon précieuse d'analyser sans risque de biais les nouveaux effets indésirables possibles, même si normalement les essais de ce genre manquent de puissance pour cette tâche.

Certains essais sont conçus dans le but de formuler des revendications de supériorité ou d'équivalence en ce qui concerne l'innocuité et la tolérabilité dans la comparaison de médicaments ou de doses du même médicament. Ces revendications doivent être appuyées par les résultats d'essais confirmatoires, comme ce doit être le cas pour les revendications concernant l'efficacité.

### 6.2 Choix des variables et collecte des données

Dans tout essai clinique, le choix des méthodes et des critères servant à évaluer l'innocuité et la tolérabilité d'un médicament dépendra d'un certain nombre de facteurs, à savoir la connaissance des effets indésirables de médicaments semblables, les résultats d'essais non cliniques et d'essais cliniques antérieurs et les effets probables des propriétés pharmacodynamiques et pharmacocinétiques du médicament à l'étude, le mode d'administration, la catégorie des sujets étudiés, et la durée de l'essai. Les résultats des tests de laboratoire en chimie clinique et en hématologie, les signes vitaux et les événements indésirables cliniques (maladies, signes et symptômes) constituent normalement l'essentiel des données sur l'innocuité et la tolérabilité. Il est particulièrement important d'enregistrer la survenance d'événements indésirables sérieux et les interruptions de traitement causées par des événements indésirables (voir les lignes directrices E2A et E3 de l'ICH).

En outre, pour la collecte des données et l'évaluation des résultats, il est recommandé d'utiliser les mêmes méthodes durant tout le programme d'essais cliniques pour pouvoir regrouper plus facilement les données d'essais différents. L'emploi d'un «dictionnaire d'événements indésirables» est particulièrement important. Ce dictionnaire est organisé de manière à présenter les données relatives aux événements indésirables selon trois niveaux d'agrégation: système-appareil-organe, terme recommandé, terme répertorié (voir le glossaire). L'agrégation des

données se fait habituellement au niveau du terme recommandé; dans la présentation descriptive des données on peut alors regrouper les termes recommandés qui appartiennent à la même classe système-appareil-organe (voir la ligne directrice M1 de l'ICH).

### 6.3 Ensemble des sujets à évaluer et présentation des données

Pour l'évaluation globale de l'innocuité et de la tolérabilité, on définit normalement l'ensemble des sujets à étudier comme les sujets ayant reçu au moins une dose du médicament de recherche. On doit recueillir le plus de données possible sur l'innocuité et la tolérabilité auprès de ce groupe de personnes (p. ex., type d'événement indésirable, gravité, apparition et durée (voir la ligne directrice E2B de l'ICH)). On peut devoir recueillir des renseignements additionnels sur l'innocuité et la tolérabilité pour des sous-groupes particuliers, comme les femmes, les aînés (voir la ligne directrice E7 de l'ICH), les personnes souffrant d'une maladie grave, ou celles qui subissent le même traitement simultanément. Ces études peuvent devoir examiner des questions plus précises (voir la ligne directrice E3 de l'ICH).

On devra porter attention à toutes les variables d'innocuité et de tolérabilité durant l'évaluation et exposer la méthodologie générale dans le protocole. On devra rapporter tous les événements indésirables, qu'ils soient ou non réputés avoir un rapport avec le traitement. L'évaluation doit tenir compte de toutes les données de la population étudiée. On doit définir avec soin les unités de mesure et les domaines de référence des variables de laboratoire; si des unités de mesure différentes ou des domaines de référence différents sont utilisés dans le même essai (p. ex., si plus d'un laboratoire participe à l'essai), il faudra normaliser convenablement les mesures pour garantir une évaluation uniforme. On doit spécifier au préalable l'usage d'une échelle de mesure de la toxicité et en donner la justification.

La fréquence d'un événement indésirable est exprimée habituellement comme le rapport du nombre de sujets ayant subi l'événement au nombre de sujets à risque. Cependant, le calcul de ce rapport ne va pas toujours de soi. Par exemple, suivant les circonstances, on peut envisager de faire figurer au dénominateur le nombre des sujets exposés ou l'ampleur de l'exposition (en années-personnes). Que le calcul ait pour but d'estimer un risque ou d'établir une comparaison entre des groupes de traitement, il est important de définir ce rapport dans le protocole, et même primordial s'il s'agit d'un traitement de longue durée et si on s'attend à observer une forte proportion d'arrêts de traitement ou de décès. Dans de tels cas, on doit envisager de recourir à des méthodes d'analyse des durées de vie et calculer des taux cumulatifs d'événements indésirables afin d'éliminer le risque de sous-estimation.

Dans les cas où il y a un bruit de fond considérable de signes et symptômes (par exemple, dans les essais psychiatriques), on doit examiner des façons d'en tenir compte dans l'estimation du risque de survenance d'événements indésirables. Une de ces façons est l'application du concept de la «suite du traitement» (voir le glossaire), selon lequel on enregistre des événements indésirables uniquement s'ils se manifestent pour la première fois après le début du traitement ou s'ils s'amplifient une fois que le traitement a débuté.

On peut aussi envisager d'autres méthodes pour réduire l'effet du bruit de fond, par exemple celle qui consiste à faire abstraction des événements indésirables de faible gravité ou la méthode qui consiste à poser la condition selon laquelle un événement devra avoir été observé de façon répétée pour être inclus dans le numérateur. Ces méthodes doivent être décrites et justifiées dans le protocole.

#### **6.4 Évaluation statistique**

L'étude de l'innocuité et de la tolérabilité est une question à plusieurs facettes. Bien que l'on puisse normalement prévoir la manifestation de certains effets indésirables des médicaments et que l'on puisse en suivre l'évolution, l'éventail des effets indésirables possibles est très large et il peut toujours en apparaître de nouveaux, dont on ne soupçonne même pas l'existence. De plus, la survenance d'un événement indésirable par suite d'un écart au protocole, comme l'usage d'un médicament non étudié, peut créer un biais. Ces conditions font ressortir les difficultés statistiques qui accompagnent l'évaluation de l'innocuité et de la tolérabilité des médicaments et signifient qu'il est plutôt exceptionnel de pouvoir tirer des conclusions claires des essais cliniques confirmatoires.

Dans la plupart des essais, on examinera avec grand profit les effets de l'innocuité et de la tolérabilité en appliquant aux données des méthodes de statistique descriptive et en calculant des intervalles de confiance lorsque cela est utile à l'interprétation des résultats. On gagnera aussi à recourir à des représentations graphiques qui décrivent le profil des événements indésirables dans les groupes de traitements et chez les sujets.

Il est parfois utile de calculer des valeurs  $p$  soit pour évaluer un écart particulier ou pour faire ressortir des écarts qui méritent une plus grande attention; dans ce dernier cas, la valeur  $p$  sert d'«indice de signalisation», que l'on applique à un grand nombre de variables d'innocuité et de tolérabilité. Ce calcul est particulièrement utile pour les données de laboratoire, qui, autrement, peuvent être difficiles à agréger convenablement. Il est recommandé de soumettre les données de laboratoire à une analyse quantitative (p. ex., évaluation des moyennes pour les traitements) et à une analyse qualitative, où on détermine le nombre de sujets qui se trouvent au-dessus ou au-dessous d'un certain seuil.

Si l'on utilise des tests d'hypothèses, il convient de faire des corrections statistiques pour multiplicité afin de quantifier l'erreur de première espèce, mais l'erreur de seconde espèce est habituellement plus pertinente dans les circonstances. On doit interpréter avec prudence les résultats réputés statistiquement significatifs lorsqu'il n'y a pas de correction pour multiplicité.

Dans la plupart des essais, les investigateurs cherchent à établir qu'il n'y a pas d'écart cliniquement inacceptable entre deux médicaments ou entre un médicament et un placebo au point de vue de l'innocuité et de la tolérabilité. Comme pour les tests de non-infériorité ou les essais d'équivalence dans l'évaluation de l'efficacité, on préfère les intervalles de confiance aux tests d'hypothèses dans les circonstances. Ainsi, on peut démontrer clairement la grande imprécision qui découle souvent des faibles fréquences.

## 6.5 Sommaire intégré

En règle générale, on présente pour les divers essais un résumé des propriétés d'innocuité et de tolérabilité du médicament de recherche tout le long de la mise au point de ce médicament et en particulier au moment d'une demande de mise en marché. Toutefois, ce sommaire sera utile dans la mesure où les essais auront été menés selon les normes et que la qualité des données sera élevée.

On doit toujours envisager l'utilité globale d'un médicament comme une question d'équilibre entre des risques et des avantages; on peut également le voir ainsi pour un seul essai, même si l'évaluation des risques-avantages est faite habituellement dans le sommaire du programme général des essais cliniques (voir section 7.2.2). Pour plus de détails sur les rapports concernant l'innocuité et la tolérabilité, voir le chapitre 12 de la ligne directrice E3 de l'ICH.

## 7. RAPPORT

### 7.1 Évaluation et rapport

Comme il a été mentionné dans l'Introduction, la ligne directrice E3 de l'ICH a pour objet la structure et le contenu des rapports d'étude clinique. Cette ligne directrice traite en détail la présentation de la méthodologie statistique, des données cliniques et des autres renseignements pertinents. La présente section est donc relativement courte.

À l'étape de la planification de l'essai, on doit inscrire dans le protocole les principales caractéristiques de l'analyse, comme il est décrit dans la section 5. Lorsque l'essai est terminé et que les données ont été rassemblées en vue de l'examen préliminaire, il est utile d'effectuer

l'examen aveugle comme il est aussi décrit dans la section 5. Cette étape, qui se fait sans la connaissance du traitement, est l'occasion de décider, par exemple, de l'exclusion de sujets ou de données des ensembles d'analyse; c'est aussi l'occasion de vérifier si des transformations sont possibles et de définir les valeurs aberrantes. On peut aussi ajouter dans le modèle des covariables importantes qui ont été définies dans d'autres études récentes; enfin, on peut reconsidérer l'usage des méthodes paramétriques ou non paramétriques. On doit inscrire dans le rapport les décisions prises à cette étape et les présenter séparément de celles prises une fois que le statisticien a obtenu les codes de traitement, car les décisions à l'aveugle réduisent généralement les risques de biais. Les statisticiens ou les autres membres du personnel qui font des analyses intermédiaires sans insu ne doivent pas participer à l'examen aveugle ni modifier le plan d'analyse statistique. Si l'insu est compromis par la possibilité que les données laissent paraître les effets du traitement, on devra prendre beaucoup de précautions pour effectuer l'examen aveugle.

Bon nombre des aspects plus pointus de la présentation et de la totalisation doivent être précisés à l'étape de l'examen aveugle, de sorte que l'on connaisse à fond tous les aspects de l'analyse avant de l'amorcer: choix des sujets, choix et modification des données, récapitulation et totalisation des données, estimation et tests d'hypothèses. Une fois la validation des données terminée, on doit exécuter l'analyse selon le plan prévu; plus on se conformera au plan, plus les résultats seront crédibles. On portera une attention particulière aux différences entre l'analyse qui était prévue et celle qui a été effectuée, selon la description du protocole, ainsi qu'aux modifications au protocole et à la mise à jour du plan d'analyse statistique fondée sur un examen aveugle des données. S'il y a lieu, on devra expliquer en détail pourquoi l'analyse qui a été effectuée diffère de celle qui était prévue.

Le rapport doit tenir compte de tous les sujets qui avaient été acceptés pour l'essai, qu'ils soient ou non inclus dans l'analyse. Tous les motifs d'exclusion de l'analyse doivent être expliqués en bonne et due forme; on doit aussi expliquer, s'il y a lieu, pourquoi un sujet faisant partie de l'ensemble d'analyse intégral n'est pas inclus dans l'ensemble conforme au protocole. De même, le rapport devra tenir compte des valeurs de chaque variable importante mesurées à toutes les périodes pertinentes pour tous les sujets faisant partie de l'ensemble d'analyse.

On doit considérer attentivement l'effet des pertes de sujets ou de données, des interruptions de traitement et des écarts majeurs au protocole sur les principales analyses de la variable primaire. On doit relever les cas de suivi impossible, d'interruption de traitement ou de non-respect grave du protocole et en faire une analyse descriptive, en donnant les raisons de ces pertes et en décrivant leur rapport avec le traitement et les résultats.

Les statistiques descriptives constituent un élément indispensable des rapports. Par des tableaux et des représentations graphiques appropriés, on doit pouvoir illustrer clairement les caractéristiques majeures des variables primaires et secondaires et des principales variables pronostiques et démographiques. Par ailleurs, on doit soigner d'une façon particulière la présentation des résultats des principales analyses ayant trait aux objectifs de l'essai. Pour la présentation des résultats de tests de signification, on doit indiquer des valeurs  $p$  précises (ex.,  $p = 0,034$ ) au lieu de se limiter à mentionner les valeurs critiques.

Bien que l'analyse des résultats d'un essai clinique ait pour objectif premier de répondre aux questions qui ont rapport aux principaux objectifs de l'essai, les données observées peuvent susciter de nouvelles questions durant l'analyse avec décodage de l'insu, ce qui pourrait entraîner de nouvelles analyses statistiques, peut-être complexes. On devra faire clairement la distinction dans le rapport entre ces analyses et celles prévues dans le protocole.

L'effet du hasard peut créer des déséquilibres imprévus entre les groupes de traitement en ce qui a trait aux mesures de base qui n'ont pas été définies au départ comme des covariables mais qui ont néanmoins une valeur pronostique. La meilleure façon de résoudre la difficulté est de montrer qu'on arrive essentiellement aux mêmes conclusions que l'analyse initiale avec une analyse qui tient compte de ces déséquilibres. Si ce n'est pas le cas, on doit examiner l'incidence des déséquilibres sur les conclusions.

En règle générale, on doit recourir modérément à l'analyse non planifiée. Celle-ci est utilisée surtout lorsqu'on croit que l'effet du traitement peut varier en fonction d'autres facteurs. On peut alors tenter de définir des sous-groupes de sujets pour lesquels l'effet du traitement est particulièrement bénéfique. Cependant, on connaît bien les risques auxquels on s'expose en faisant une «sur-interprétation» des résultats des analyses de sous-groupes non planifiées (voir aussi section 5.7), et il faut donc être prudent. Bien que des problèmes d'interprétation puissent aussi survenir lorsqu'un traitement semble n'avoir aucun effet bénéfique ou qu'il semble avoir un effet indésirable dans un sous-groupe de sujets, on doit examiner adéquatement cette éventualité et en faire état dans le rapport.

Enfin, l'analyse, l'interprétation et la présentation des résultats d'un essai clinique devront être empreintes de jugement statistique. À cette fin, le statisticien affecté à l'essai devra faire partie de l'équipe chargée de la rédaction du rapport d'étude clinique et il devra en outre approuver le rapport.

## 7.2 Sommaire de la base de données cliniques

Pour pouvoir faire une demande de mise en marché, on doit produire un résumé des données relatives à l'efficacité et à l'innocuité tirées de tous les essais cliniques pertinents (ce résumé est appelé rapport d'expert dans l'Union européenne, «integrated summary» aux États-Unis et «Gaiyou» au Japon). Ce sommaire peut, le cas échéant, être accompagné d'un rapport statistique.

Le résumé comprend un certain nombre de points de nature proprement statistique: description des caractéristiques démographiques et cliniques du groupe des personnes traitées dans le cadre du programme d'essais cliniques; étude des questions fondamentales de l'efficacité par l'examen des résultats des essais pertinents (essais avec témoins pour la plupart) et confrontation de ces résultats pour vérifier dans quelle mesure ils se complètent ou se contredisent; présentation sommaire des données sur l'innocuité contenues dans la base de données combinée des essais menés en vue de la demande de mise en marché, et définition de problèmes relatifs à l'innocuité. Durant l'élaboration d'un programme d'essais cliniques, on prendra soin de définir les mesures et les méthodes de collecte d'une manière uniforme dans le but de faciliter l'interprétation des séries d'essais, surtout si l'on regroupe les données de divers essais. On se servira d'un ouvrage de référence courant pour enregistrer les particularités des médicaments, les antécédents médicaux et les événements indésirables. Il est presque toujours utile, et essentiel pour la méta-analyse, d'avoir une définition courante des variables primaire et secondaire. Enfin, on devra garantir la compatibilité des procédures (mesure des principales variables d'efficacité, détermination du moment des évaluations par rapport à la randomisation ou à l'entrée des sujets, traitement des écarts au protocole et de leurs causes et, peut-être, définition des facteurs pronostiques), à moins que l'on ait des motifs valables pour ne pas le faire.

On doit décrire en détail les méthodes statistiques utilisées pour combiner les données de différents essais. On devra par ailleurs considérer la possibilité d'un biais découlant du choix des essais, l'homogénéité des résultats de ces essais et la manière appropriée de modéliser les diverses sources de variation. On devra aussi examiner dans quelle mesure les hypothèses et les choix effectués influent sur les conclusions des essais.

### 7.2.1 Données relatives à l'efficacité

L'essai clinique doit toujours être suffisamment important pour pouvoir remplir ses objectifs. On pourra par ailleurs obtenir d'autres données utiles en dressant le sommaire d'une série d'essais cliniques qui visent à répondre essentiellement aux mêmes grandes questions relatives à l'efficacité. Les principaux résultats de cette série d'essais devront

tous être présentés sous la même forme pour faciliter la comparaison (habituellement sous forme de tableaux ou de graphiques ayant principalement pour objet des estimations et des limites de confiance). On aura souvent intérêt à utiliser des méthodes de méta-analyse pour regrouper les estimations, parce qu'on peut ainsi estimer avec plus de précision la taille des effets du traitement, et produire un résumé complet et concis des résultats des essais. Dans des cas exceptionnels, la méta-analyse pourra être aussi la façon la plus convenable, sinon la seule façon, de démontrer l'efficacité d'un traitement au moyen d'un test d'hypothèse global. Dans ce cas particulier, on aura pris soin de rédiger à l'avance un protocole indépendant pour la méta-analyse.

### **7.2.2 Données relatives à l'innocuité**

Lorsqu'on résume les données relatives à l'innocuité, il importe de scruter la base de données pertinente afin d'y déceler des indices de toxicité et, le cas échéant, d'approfondir ces indices en cherchant la confirmation dans les tendances des observations. L'ensemble des données sur l'innocuité touchant tous les sujets ayant expérimenté un médicament est une source d'information importante, parce qu'en raison de sa taille il crée les meilleures conditions pour déceler les événements indésirables plus rares et, peut-être, en estimer la fréquence. Or, cette base de données permet difficilement d'évaluer cette fréquence à cause de l'absence de groupes de comparaison; les données d'essais comparatifs sont donc particulièrement utiles pour surmonter cette difficulté. On regroupera les résultats des essais qui utilisent un comparateur commun (placebo ou comparateur actif spécifié) et on les présentera séparément pour chaque comparateur qui produit suffisamment de données.

On doit faire état de tous les indices de toxicité potentielle que révèle l'examen des données. En évaluant la réalité de ces effets indésirables, on tiendra compte de la multiplicité engendrée par l'existence de nombreuses comparaisons; par la même occasion, on utilisera convenablement les méthodes d'analyse des durées de vie pour décrire la relation probable entre la fréquence des événements indésirables et la durée du traitement ou du suivi. On devra mesurer convenablement les risques associés aux effets indésirables pour pouvoir évaluer avec justesse le rapport risques-avantages.



## GLOSSAIRE

### **Abandon**

Dans un essai clinique, sujet qui, pour une raison quelconque, décide de se retirer de l'expérience avant la date prévue par le protocole.

### **Analyse intermédiaire**

Analyse qui a pour but de comparer des groupes d'essai sur le plan de l'efficacité ou de l'innocuité à n'importe quel moment avant la conclusion de l'essai.

### **Biais (statistique et opérationnel)**

Tendance systématique d'un facteur lié à l'élaboration, à la conduite et à l'analyse d'un essai clinique, et à l'évaluation de ses résultats, à faire dévier l'estimation de l'effet d'un traitement par rapport à la valeur vraie. Le biais introduit durant la conduite de l'essai est appelé «biais opérationnel», tandis que celui introduit durant les autres étapes mentionnées ci-dessus est appelé «biais statistique».

### **Comité indépendant de surveillance des données (CISD) (comité de surveillance de l'innocuité et des données, comité de surveillance, comité de surveillance des données)**

Comité indépendant de contrôle des données pouvant être établi par le promoteur et chargé, d'une part, d'évaluer périodiquement le déroulement d'un essai clinique, les données sur l'innocuité et les résultats critiques concernant l'efficacité et, d'autre part, de formuler des recommandations concernant la poursuite, la modification ou l'interruption d'un essai.

### **Double placebo**

Technique de maintien de l'insu appliquée durant l'administration de médicaments dans un essai clinique où les deux traitements ne peuvent être identiques. Des médicaments sont préparés pour le traitement A (substance active et placebo indifférenciés) et le traitement B (substance active et placebo indifférenciés). Les sujets se soumettent alors à deux séries de traitement: soit A (substance active) et B (placebo), ou A (placebo) et B (substance active).

### **Effet du traitement**

Effet imputé à un traitement au cours d'un essai clinique. Dans la plupart des essais cliniques, l'effet étudié est observé par suite de la comparaison de deux traitements ou plus.

## **Ensemble d'analyse intégral**

Ensemble des sujets qui se rapproche le plus de l'ensemble idéal associé au principe de l'intention de traiter. Il correspond à l'ensemble des sujets randomisés, dont ont été exclus un nombre minimum de sujets non sans raison valable.

## **Ensemble conforme au protocole (cas valides, échantillon d'efficacité, échantillon des sujets évaluable)**

Ensemble des données provenant du sous-ensemble des sujets qui se sont conformés au protocole d'une manière suffisamment rigoureuse pour que l'on puisse affirmer que ces données révèlent les effets du traitement tels que le laissait prévoir le modèle scientifique utilisé. Le respect du protocole est mesuré à l'aide de critères tels que l'exposition au traitement, l'existence d'observations et l'absence d'écarts majeurs au protocole.

## **Essai d'équivalence**

Essai dont l'objectif principal est de montrer qu'il existe une différence cliniquement négligeable entre les réactions à deux traitements ou plus; en règle générale, la différence réelle entre des traitements doit se situer dans un intervalle «d'équivalence» ayant une limite inférieure et une limite supérieure, défini pour les différences cliniquement acceptables.

## **Essai multicentrique**

Essai clinique mené selon un même protocole à plusieurs endroits en même temps et donc, réalisé par plus d'un investigateur.

## **Examen aveugle**

Opération qui consiste à vérifier et à évaluer les données d'un essai entre le moment où celui-ci prend fin (la dernière observation faite sur le dernier sujet) et le moment où l'insu est décodé, dans le but de mettre au point les derniers détails de l'analyse prévue.

## **Fiabilité inter-évaluateur**

Propriété d'une variable qui consiste à produire des résultats équivalents lorsqu'elle est utilisée par différents évaluateurs dans des situations différentes.

## **Fiabilité intra-évaluateur**

Propriété d'une variable qui consiste à produire des résultats équivalents lorsqu'elle est utilisée par le même évaluateur dans des situations différentes.

## **Généralisation**

Capacité d'étendre avec fiabilité les résultats d'un essai clinique mené avec un échantillon de sujets à une population de patients et à un éventail plus large de conditions cliniques.

## **Innocuité et tolérabilité**

L'innocuité d'un produit médical a trait au risque médical auquel est exposé le sujet; dans un essai clinique, on mesure habituellement ce risque par des tests de laboratoire (tests en chimie clinique et en hématologie), les signes vitaux, les événements indésirables cliniques (maladies, signes et symptômes) et d'autres épreuves spéciales (p. ex., ECG, ophtalmologie). La tolérabilité du produit médical désigne le degré d'effets indésirables manifestes du produit que peut tolérer le sujet.

## **Interaction (qualitative et quantitative)**

Il y a interaction lorsque la divergence de traitement (p. ex., différence entre le produit à l'étude et la substance-témoin) dépend d'un facteur tiers (p. ex., centre d'essai); il y a interaction quantitative lorsque le degré de divergence varie selon la valeur du facteur, et il y a interaction qualitative lorsque le sens de la divergence varie pour au moins une valeur du facteur.

## **Méta-analyse**

Évaluation formelle des données quantitatives de deux essais ou plus qui visent à répondre à la même question. Cette opération consiste le plus souvent à regrouper les statistiques sommaires des divers essais, mais le terme est utilisé parfois pour désigner le regroupement des données brutes.

## **Méthodes axées sur l'observation de fréquences**

Méthodes statistiques, telles que les tests de signification et la construction d'intervalles de confiance, dont l'objet est la fréquence des événements qui surviennent dans des réalisations successives hypothétiques de la même expérience.

## Méthodes bayésiennes

Méthodes d'analyse des données qui définissent une distribution de probabilité *a posteriori*, fondée sur les observations, pour un paramètre quelconque (p. ex., effet du traitement), ainsi qu'une distribution de probabilité *a priori* pour ce paramètre, la première servant alors de base à l'inférence statistique.

## Plan d'analyse statistique

Document contenant une description plus détaillée et plus technique des principaux aspects de l'analyse contenus dans le protocole et exposant les modalités de l'analyse statistique appliquée aux variables primaires et secondaires et aux autres données.

## Principe de l'intention de traiter

Principe suivant lequel on peut le mieux évaluer l'effet d'un programme de traitement en fondant cette évaluation sur l'intention de traiter un sujet (c.-à-d., application du régime de traitement prévu) plutôt que sur le traitement réellement administré. Une des conséquences de ce principe est que les sujets affectés à un groupe de traitement seront suivis, évalués et analysés en tant que membres de ce groupe, qu'ils se conforment ou non à la série de traitement prévue.

## Statisticien affecté à l'essai

Statisticien qui possède la formation et l'expérience nécessaires pour appliquer les principes contenus dans la présente ligne directrice et qui est responsable des aspects statistiques de l'essai.

## Suite du traitement

Événement qui survient durant le traitement et qui marque soit un changement imprévu dans l'état du sujet, soit une détérioration de l'état du sujet par rapport à la période d'avant traitement.

## Termes recommandés et termes répertoriés

Dans un dictionnaire médical avec classement hiérarchique (p. ex., MedDRA), les termes répertoriés constituent le dernier niveau de définition, celui auquel se réfère le chercheur pour le codage des descriptions, tandis que les termes recommandés constituent le niveau d'agrégation des termes répertoriés qui sert habituellement à l'enregistrement de la fréquence des événements. Par exemple, la description du chercheur «douleur au bras gauche» pourrait être classée sous «douleur articulaire» (terme répertorié), appellation qui figure sous «arthralgie» au niveau des termes recommandés.

### **Test de non-infériorité**

Essai dont l'objectif principal est de montrer que la réaction provoquée par le produit de recherche n'est pas cliniquement inférieure à celle provoquée par un comparateur (substance active ou placebo comme témoin).

### **Test de supériorité**

Essai dont l'objectif principal est de montrer que la réaction provoquée par le produit de recherche est supérieure à celle provoquée par un comparateur (substance active ou placebo comme témoin).

### **Validité de contenu**

Capacité d'une variable (p. ex., échelle d'évaluation) de mesurer ce qu'elle est censée mesurer.

### **Variable d'évaluation globale**

Variable unique (en règle générale, une échelle de valeurs nominales ordonnées) qui intègre des variables objectives et l'opinion générale du chercheur sur l'état du sujet ou l'évolution de cet état.

### **Variable de substitution**

Variable qui sert à mesurer indirectement un effet clinique lorsqu'il n'est pas possible de mesurer directement cet effet.