

とを実証することを目的とした実対照試験では、特に有用である。ベイズ法を用いた試験ではこの問題を過去の試験を用いて実対照と無治療対照の比較を行うことによって検討することができる。ベイズ法の実対照試験への適用についてはGould (1991) およびSimon (1999) の考察を参照のこと。

5.5 Initial information about the endpoints: prior distributions

5.5 エンドポイントに関する最初の情報：事前分布

The initial uncertainty about the endpoints or parameters of interest, both in the control and treatment groups, is quantified through probability distributions, called prior distributions. See Gelman et al. (2004) for background on different types of prior distributions. You should select the appropriate prior information and incorporate it into the analysis correctly. Discussions regarding study design with FDA will include an evaluation of the model to be used to incorporate the prior information into the analysis. Irony & Pennello (2001) discuss prior distributions for trials under regulatory review.

対照群および治療群の両群において、関心のあるエンドポイントまたはパラメータに関する最初の不確実性は、事前分布と呼ばれる確率分布によって定量化される。様々なタイプの前分布に関する背景についてはGelman et al. (2004) を参照のこと。適切な事前知識を選択し、それを正しく解析に組み込まなくてはならない。試験デザインについてのFDAとの検討には、事前知識の解析への組み込みに用いるモデルの評価も含む。Irony & Pennello (2001) は規制検討下の試験の前分布について考察している。

Informative priors

情報事前分布

Prior knowledge is described by an informative prior distribution. Because using prior information may decrease the sample size in a trial, we recommend you identify as many sources of good prior information as possible when planning a trial. FDA should agree with your choice of prior distributions. Possible sources of prior information include:

- clinical trials conducted overseas
- patient registries
- clinical data on very similar products
- pilot studies.

事前知識は情報事前分布によって示される。事前知識を用いることによって臨床試験におけるサンプルサイズが縮小できるため、臨床試験の計画時には、可能な限り多くの良好な事前知識源を同定することを推奨する。FDAは選択された事前分布について合意するものとする。事前知識源となり得るものは、

以下の通りである。

- 海外で実施された臨床試験
- 患者登録簿
- よく類似した製品に関する臨床データ
- パイロット試験

We recommend the proposed prior information be submitted as part of the IDE (when an IDE is required). In some cases, existing valid prior information may be unavailable for legal or other reasons (e.g., the data may belong to someone else who is unwilling to allow access).

上記に提示された事前知識は（IDE申請が必要な場合は）IDE申請の一部として提出することを推奨する。場合によっては、既存の有効な事前知識が法規制等の理由から使用できない場合がある（例：当該データが閲覧許可に積極的ではない他者に属する場合）。

We recommend you hold a pre-IDE meeting with FDA to come to agreement on what prior information is scientifically valid and how it will be used in the analysis. Quantitative priors (i.e., those based on data from other studies) are the easiest to evaluate. We recommend the prior studies be similar to the current study in as many as possible of the following aspects:

- protocol
- endpoints
- target population
- sites
- physicians
- time frame.

FDAとIDE申請前に検討会を行い、どの事前知識が科学的に妥当でありそれをどのように解析に用いるかについて合意しておくことを推奨する。定量的な事前知識（すなわち他の試験から得られたデータに基づくもの）は最も評価が容易である。過去の試験は以下の点で可能な限り現在の試験と類似していることが望ましい。

- 治験実施計画書
- エンドポイント
- 標的母集団
- 試験実施施設
- 試験医師
- 試験期間

Priors based on expert opinion rather than data are problematic. Approval of a device could be delayed or jeopardized if FDA advisory panel members or other clinical evaluators do not agree with the opinions used to generate the prior. データよりも専門家の意見に基づいた事前知識は問題がある。FDA諮問委員会のメンバーやその他の臨床評価担当者がそのような事前知識の作成に用いられた意見に合意しない場合、医療機器の承認が遅れたり脅かされたりする可能性がある。

To avoid bias, we recommend you avoid using studies that are not representative (e.g., if non-favorable studies are systematically excluded). We recommend you check for selection bias by examining:

- the population of all studies that might have been considered
- the reasons for including or excluding each study.

バイアスを回避するため、典型的でない試験データを用いることは避けるよう推奨する（例：好ましくない試験が体系的に除外される場合）。以下を確認することにより選択バイアスを確認することを推奨する。

- 検討されたすべての試験の母集団
- 各試験の組み入れ・除外の理由

A Bayesian analysis of a current study of a new device may include prior information from:

- the new device
- the control device
- both devices.

新たな機器に関する現在の試験のベイズ解析には、以下から得た事前知識が含まれる。

- 新たな機器
- 対照機器
- 上記の両方

Most commonly, prior information based on historical controls is “borrowed,” which can significantly decrease the sample size in a concurrent control group. As a result, a greater proportion of patients can be allocated to the experimental device, increasing the experience with that device at a faster pace. However, if differences between the historical control studies and the current study are large, the use of a historical control as prior information for a concurrent control may not be advantageous.

通常、歴史的対照に基づく事前知識は「借り物」であり、同時対照群のサン

プルサイズを有意に縮小することができる。この結果、より多くの患者を試験対象機器に割り付けることができ、当該機器による実験量がより速いペースで増加する。しかし、歴史的対照試験と現在の試験との差異が大きい場合、歴史的対照を同時対照の事前知識として用いることは有用ではないこともある。

For example, consider a study with the objective of demonstrating the experimental device is non-inferior to the control regarding rate of complication. If patients in the study are more likely to have complications than those in the historical control studies because they are sicker, for example, indiscriminate use of the historical control data will bias downward the concurrent control complication rate. This bias will make it more difficult to demonstrate the experimental device is non-inferior to the control than if the historical control data were ignored. This phenomenon is less likely to occur if the historical control data are properly calibrated to the current study by, for example, adjusting for important covariates.

試験対象機器が、合併症の発生率に関して対照に劣らないことを実証することを目的とした試験を例に挙げて考察することにする。例えば被験者の病状悪化のため、歴史的対照試験の場合より合併症の発生率が高い傾向にある場合、歴史的対照のデータの無差別な使用によって同時対照の合併症発生率が減少するというバイアスがかかる。このバイアスにより、歴史的対照のデータを無視した場合よりも試験対象機器が対照に劣っていないことを実証することがさらに困難となる。重要な共変量の調整などによって歴史的対照のデータを現在の試験にあわせて適切に校正すれば、この現象が発生することは少ない。

If the prior information for a study is based on many more patients than are to be enrolled in the study, the prior distribution may be too informative. In this case, the prior probability that the pivotal study is a success (i.e., demonstrates the proposed claims) will be excessively high. If the prior probability of a successful study is too high, it can be lowered in various ways, including:

- modifying or discounting the prior distribution in some way
- increasing the stringency of the decision rule.

ある試験の事前知識が当該試験に登録される患者より多くの患者に基づくものである場合、事前分布は非常に有益となる場合がある。この場合、主試験が成功する事前確率（すなわち提示された主張が実証される確率）は過度に高くなる。試験成功の事前確率が非常に高い場合は、以下を含む様々な方法でその確率を低下させることができる。

- 何らかの方法で事前分布の修正または割引を行う
- 決定則の厳密性を増大させる

Non-informative priors

無情報事前分布

Lack of any prior knowledge may be reflected by a non-informative prior distribution. Usually, it is easy to define maximum and minimum values for the parameters of interest, and in this case, a possible “non-informative” prior distribution is a uniform distribution ranging from the minimum to the maximum value.

事前知識の欠落は無情報分布によって反映される。通常、関心のあるパラメータの最大値および最小値の定義は容易であり、この場合、考えられる「無情報」事前分布は最小値から最大値までの一様分布である。

Non-informative priors are reviewed in Kass & Wasserman (1996). Standard and related improper priors are explained and used extensively in Box & Tiao (1973). Reference priors are extensively discussed in Bernardo & Smith (1993).

無情報事前分布についてはKass & Wasserman (1996) が検証している。標準的な事前分布および関連する不適切な事前分布については、Box & Tiao (1973) による文献で解説され広範にわたって用いられている。Reference priorと呼ばれる事前分布についてはBernardo & Smith (1993) が広範にわたり考察している。

5.6 Borrowing strength from other studies: hierarchical models

5.6 他の試験からの説得力の借用：階層モデル

Bayesian hierarchical modeling is a specific methodology you may use to combine prior results with a current study to obtain estimates of safety and effectiveness parameters. The name hierarchical model derives from the hierarchy in which observations and parameters are structured. The Bayesian analyst refers to this approach as “borrowing strength.” For device trials, strength can be translated into sample size, and the extent of borrowing depends on how closely results from the new study reflect the prior experience.

ベイズ法による階層モデリングは、既知の結果を現在の試験結果と結合して安全性および有効性のパラメータを推定するために用いることのできる特殊な手法である。「階層モデル」という名称は、観察事項およびパラメータの構築における階層性から派生したものである。ベイズ解析ではこの手法を「説得力の借用 (borrowing strength)」と称している。医療機器の臨床試験では、説得力 (strength) はサンプルサイズに転換され、借用の程度は新たな試験の結果がどの程度緊密に事前知識を反映しているかによって異なる。

If results are very similar, the current study can borrow great strength. As current results vary from the previous information, the current study borrows less and less. Very different results borrow no strength at all, or even potentially “borrow

negatively”. In a regulatory setting, hierarchical models can be very appealing: They reward having good prior information on device performance by lessening the burden in demonstrating safety and effectiveness. At the same time, the approach protects against over-reliance on previous studies that turn out to be overly optimistic for the pivotal study parameter.

結果が非常に類似している場合、現在の試験は大きな説得力を借用することができる。現在の試験の結果が事前知識と異なれば、現在の試験が借用する説得力は少なくなる。結果が大きく異なる場合には、説得力はまったく借用されないか、「否定的に借用」される可能性さえある。規制設定において階層モデルは非常に魅力的である。これは、階層モデルでは安全性および有効性の実証における負担が抑えられ、それによって機器の性能に関する良好な事前知識が得られるためである。同時に、階層モデルの使用により、主試験のパラメータについて過度に楽観的であると判明した過去の試験に関する過剰な信頼を避けることができる。

An example hierarchical model

階層モデルの例

Suppose you want to combine information from a treatment registry of an approved device with results from a new study. You may decide to use two hierarchical levels: the patient level and the study level.

承認済みの機器の治療記録から得られた情報を新たな試験の結果と結合する場合について考察する。この場合、患者レベルおよび試験レベルという2つの階層レベルの使用が考えられる。

The first (patient) level of the hierarchy assumes that (1) within the current study, patients are exchangeable; and (2) within the registry, patients are exchangeable. Registry patients are not, however, exchangeable with patients in the current study, so patient data from the registry and the current study may not be simply pooled. 1つめの階層レベル（患者）では、（1）現在の試験において患者が交換可能である、（2）治療記録において患者が交換可能である、と仮定する。しかし、治療記録の患者は現在の試験の患者と交換することはできないため、治療記録と現在の試験から得られた患者データは単にプールすることはできない。

The second (study) level of the hierarchy applies a model that assumes the success probabilities from the registry and the current study are exchangeable, but the rates may differ (e.g., they may depend on covariates). This assumption is prudent any time you are not sure if patients from the prior experience (i.e., the registry) are directly exchangeable with the patients from the current study. However, the two success probabilities are related in that they are assumed exchangeable. As a result, the registry provides some information about the success probability in the current

study, although not as much information as if the patients in the two groups were directly poolable.

2つめの階層レベル（試験）では、治療記録から得られた成功確率と現在の試験から得られた成功確率が交換可能であると仮定するモデルを適用するが、これらの確率には差がある（例：共変量により異なる）。この仮定は、過去の試験（すなわち治療記録）の患者が現在の試験の患者と直接交換可能であるかどうか不明である場合には賢明である。しかし、この2つの成功確率は、それらが交換可能と仮定されているという点で関連性がある。このため、2つの群の患者が直接プールできる場合ほど多くの情報ではないにせよ、治療記録から、現在の試験の成功確率について何らかの情報を得ることができる。

Similarity of previous studies to current study

過去の試験と現在の試験との類似性

The key clinical question in using hierarchical modeling to borrow strength from previous studies is whether the previous studies are sufficiently similar to the current study in covariates such as:

- devices used
- objectives
- endpoints
- protocol
- patient population
- sites
- physician training
- patient management
- time frame.

過去の試験から説得力を借用するための階層モデリングの使用において重要な臨床学的問題は、過去の試験が以下の共変量において現在の試験と十分に類似性があるかどうかである。

- 使用する機器
- 目的
- エンドポイント
- 試験実施計画
- 患者母集団

- 試験実施施設
- 医師の訓練度
- 患者の管理
- 試験期間

Statistical adjustments for certain differences in covariates such as demographic and prognostic variables may be appropriate, using patient-level data. Generally, proper calibration of your study depends on using the same covariate information at the patient level as in previous studies.

人口統計学的変数や予後に関する変数などの共変量における一定の差異の統計学的調整は、患者レベルのデータを用いて適切に行うことができる。一般に、試験の適切な較正は、過去の試験と同様の患者レベルの共変量情報の使用に左右される。

Calibration based only on covariate summaries (such as from the literature) may be inadequate because the relationship of the covariate level to the outcome can be determined in the current study but not in the previous studies. This forces the untestable assumption that covariate effects in your study and previous studies are the same; that is, that study and covariate effects do not interact.

共変量の概要(参考文献など)のみに基づいた較正は適切ではない。これは、共変量のレベルと転帰との関係が過去の試験ではなく現在の試験において決定付けられるためである。これによって、現在の試験および過去の試験における共変量の影響は同じである。つまり、試験と共変量の影響とは相互作用しない、という検証不可能な仮定が強制的に生じる。

When you use more than one study as prior information in a hierarchical model, the prior distribution can be very informative. As discussed previously, if the prior probability of a successful trial is too high, we recommend the study design and analysis plan be modified.

階層モデルにおいて複数の事前知識を用いる場合、事前分布は非常に有益となり得る。前述のとおり、試験成功の事前確率が非常に高い場合は試験デザインおよび解析計画に修正を加えることを推奨する。

You may also use hierarchical models to combine data across centers in a multi-center trial. For an example, see the Summary of Safety and Effectiveness for PMA P980048, BAK/Cervical Interbody Fusion System by Sulzer Spine-Tech.⁶ 多施設共同試験において施設間のデータを結合するのに階層モデルを用いることもできる。例としては、PMA P980048 Sulzer Spine-Tech製BAK/Cervical Interbody Fusion Systemの安全性および有効性に関する概要 (Summary of

Safety and Effectiveness for PMA P980048, BAK/Cervical Interbody Fusion System by Sulzer Spine-Tech) を参照のこと。⁶

Outcomes for devices can vary substantially by site due to such differences as:

- physician training
- technique
- experience with the device
- patient management
- patient population.

機器の試験結果は試験実施施設によって大きく異なる場合がある。これは施設によって以下のような差異があるためである。

- 医師の訓練度
- 技術
- 機器使用実績
- 患者の管理
- 患者母集団

A hierarchical model on centers assumes that the parameters of interest vary from center to center but are related via exchangeability. This kind of model adjusts for center-to-center variability when estimating parameters across studies.

各試験実施施設の階層モデルでは、関心のあるパラメータは施設によって異なるが、交換可能性において関連性があると仮定する。こうしたモデルによって試験間のパラメータを推定する際に、施設間のばらつきが調整される。

Non-technical discussion of hierarchical models and technical details on their implementation appear in Gelman et al. (2004). Other, more complex approaches are described in Ibrahim & Chen (2000) and Dey et al., 1998.

階層モデルの非技術的考察およびその実施に関する技術的詳細はGelman et al. (2004) を参照のこと。その他、より複雑な手法についてはIbrahim & Chen (2000) およびDey et al. (1998) を参照のこと。

⁶ www.fda.gov/cdrh/pdf/p980048b.pdf

⁶ www.fda.gov/cdrh/pdf/p980048b.pdf

5.7 Determining the sample size

5.7 サンプルサイズの決定

The sample size in a clinical trial depends on:

- prior information
- mathematical model used in analysis
- distributions of parameters in the analytic model
- specific decision criteria
- variability of the sample.

臨床試験のサンプルサイズは、以下の条件によって異なる。

- 事前知識
- 解析に用いられる数学的モデル
- 解析モデルにおけるパラメータ分布
- 特定の決定基準
- サンプルのばらつき

If the population of patients is highly variable, the sample size increases. If there is no variability (i.e., everyone in the population has the same value for the measurement of interest), a single observation is sufficient. The purpose of sizing a trial is to gather enough information to make a decision while not wasting resources or putting patients at unnecessary risk.

患者集団のばらつきが非常に大きい場合、サンプルサイズは拡大される。ばらつきがない場合（すなわち当該集団の患者全員で関心のある観察事項の値が同じ場合）は、1つの観察を行えば十分である。臨床試験のサンプルサイズ決定の目的は、情報源を無駄にしたり患者に不要なリスクを負わせたりせずに、十分な情報を収集し決定を行うことである。

In traditional frequentist clinical trial design, the sample size is determined in advance. Instead of specifying a particular sample size, the Bayesian approach (and some modern frequentist methods) may specify a particular criterion to stop the trial. Appropriate stopping criteria may be based on a specific amount of information about the parameter (e.g., a sufficiently narrow credible interval, defined in **Section 6: Analyzing a Bayesian Clinical Trial**) or an appropriately high probability for a pre-specified hypothesis.

従来の頻度論的な臨床試験デザインでは、サンプルサイズは事前に決定される。ベイズ法（および一部の最新の頻度論的手法）では、個々のサンプルサイズを指定する代わりに、試験を中止するための特定の基準を指定することができる。適切な試験中止基準はパラメータに関する特定量の情報（例：6ベイズ法を用いた臨床試験の解析で定義される十分に狭い信用区間）または

事前に指定した仮説に関する十分高い確率に基づいて決定できる。

At any point before or during a Bayesian clinical trial, you can obtain the posterior distribution for the sample size. Therefore, at any point in the trial, you can compute the expected additional number of observations needed to meet the stopping criterion. In other words, the sample size distribution is continuously updated as the trial goes on. Because the sample size is not explicitly part of the stopping criterion, the trial can be ended at the precise point where enough information has been gathered to answer the important questions.

ベイズ法を用いた臨床試験の実施前または実施中のどの時点でも、サンプルサイズの事後分布を得ることができる。したがって、試験におけるある時点で、試験中止基準を満たすのに必要な観察事項の予測追加数を算出することができる。言い換えると、サンプルサイズの分布は試験の進行に従って継続的に更新される。サンプルサイズは試験中止基準の明確な一部ではないため、試験は重要な疑問に対する回答を得るのに十分な情報が収集された時点で終了することができる。

Special considerations when sizing a Bayesian trial

ベイズ法による試験のサンプルサイズ決定時の注意事項

When sizing a Bayesian trial, FDA recommends you decide in advance on the minimum sample size according to safety and effectiveness endpoints because safety endpoints may lead to a larger sample size. FDA also recommends you include a minimum level of information from the current trial to enable verification of model assumptions and appropriateness of prior information used. This practice also enables the clinical community to gain experience with the device.

ベイズ法による試験のサンプルサイズ決定時には、FDAは事前に安全性および有効性のエンドポイントに従って、最小限のサンプルサイズを決定しておくことを推奨している。これは、安全性のエンドポイントによってサンプルサイズが拡大されることがあるためである。また、FDAでは、モデル仮定の検証と使用する事前知識の適切性の確認が可能となるよう、最新の試験から得られた最小限のレベルの情報を盛り込むことも推奨している。この実行により、臨床の場では機器の使用実績を増加させることもできる。

When hierarchical models are used, we recommend you provide a minimum sample size for determining the amount of information that will be “borrowed” from other studies.

階層モデルを使用する場合は、他の試験から「借用」する情報量を決定するため、最小限のサンプルサイズを提示することを推奨する。

We recommend the maximum sample size be defined according to economical, ethical, and regulatory considerations.

最大限のサンプルサイズは、経済、倫理、および規制面での検討によって定めることを推奨する。

Various approaches to sizing a Bayesian trial are described in Inoue et al. (2005), Katsis & Toman (1999), Rubin & Stern (1998), Lindley (1997), and Joseph et al. (1995a,b).

ベイズ法による試験のサンプルサイズ決定のための多様な方法については、Inoue et al. (2005)、Katsis & Toman (1999)、Rubin & Stern (1998)、Lindley (1997) および Joseph et al. (1995a,b) を参照のこと。

5.8 Assessing the operating characteristics of a Bayesian design

5.8 ベイズ法による試験デザインの性能評価

Because of the inherent flexibility in the design of a Bayesian clinical trial, a thorough evaluation of the operating characteristics should be part of the trial design. This includes evaluation of:

- probability of erroneously approving an ineffective or unsafe device (type I error)
- probability of erroneously disapproving a safe and effective device (type II error)
- power (the converse of type II error: the probability of appropriately approving a safe and effective device)
- sample size distribution (and expected sample size)
- prior probability of claims for the device
- if applicable, probability of stopping at each interim look.

A more thorough discussion appears in the **Appendix**.

ベイズ法を用いた臨床試験デザインには固有の柔軟性があるため、その性能の徹底的評価が試験デザインの一部となる。これには以下の評価が含まれる。

- 有効性および安全性に欠ける機器を誤って承認する確率（第1種の過誤）
- 安全かつ有効な機器を誤って承認しない確率（第2種の過誤）
- 検出力（第2種の過誤の反対、すなわち安全かつ有効な機器を適切に承認する確率）
- サンプルサイズの分布（および予測されるサンプルサイズ）
- 機器に関する主張の事前確率
- 妥当な場合、各中間調査で試験を中止する確率

より詳細な考察については付録参照。

6. Analyzing a Bayesian Clinical Trial

6. ベイズ法を用いた臨床試験の解析

6.1 Summaries of the posterior distribution

6.1 事後分布のまとめ

The results, conclusions, and interpretation of a Bayesian analysis all rely on the posterior distribution, which contains all information from the prior distribution, combined with the results from the trial via the likelihood. Consequently, results and conclusions for a Bayesian trial are based only on the posterior distribution. FDA recommends you summarize the posterior distribution with a few numbers (e.g., posterior mean and standard deviation), especially when there are numerous endpoints to consider. FDA also recommends you include graphic representations of the appropriate distributions.

ベイズ解析の結果、結論、および解釈はすべて事後分布に依存している。事後分布には事前分布から得られたすべての情報が含まれるが、それらは尤度によって試験結果と結び付けられている。このため、ベイズ法を用いた試験の結果および結論は事後分布にのみ基づいている。FDAでは、考慮すべきエンドポイントが多数ある場合は特に、いくつかの数値（例：事後平均および標準偏差）を用いて事後分布をまとめることを推奨している。また、FDAでは、適切な分布図を示すことも推奨している。

6.2 Hypothesis testing

6.2 仮説検定

Statistical inference may include hypothesis testing or interval estimation, or both. FDA often bases approval on demonstrating claims via hypothesis tests. For Bayesian hypothesis testing, you may use the posterior distribution to calculate the probability that a particular hypothesis, either null or alternative, is true, given the observed data.

統計学的推論には、仮説検定または区間推定、もしくはその両方が含まれる。FDAではしばしば仮説検定によって主張を証明することにより承認を行う。ベイズ法による仮説検定では、事後分布を用いて、観察データをもとに特定の仮説（帰無仮説または代用仮説）が正しいものであるという確率を算出することができる。

Although probabilities of type I and II errors are frequentist notions, and not formally part of Bayesian hypothesis testing, Bayesian hypothesis tests often enjoy good frequentist properties. FDA recommends you provide the type I and II error rates of your proposed hypothesis test.

第1種および第2種の過誤の確率は頻度論的概念であり、正式にはベイズ法による仮説検定の一部ではないが、ベイズ法による仮説検定では良好な頻度論的特性がみられることが多い。FDAは、提示する仮説検定について第1種および第2種の過誤の確率を示すことを推奨している。

6.3 Interval estimation

6.3 区間推定

Bayesian interval estimates are based on the posterior distribution and are called credible intervals. If the posterior probability that an endpoint lies in an interval is 0.95, then this interval is called a 95 percent credible interval.

ベイズ法による区間推定は事後分布に基づいて行われるが、これは信用区間と呼ばれる。エンドポイントがある区間に位置するという事後確率が0.95である場合、この区間を95%信用区間という。

For construction of credible intervals, see Chen & Shao (1999) and Irony (1992). Other types of Bayesian statistical intervals include highest posterior density (HPD) intervals (Lee, 1997) and central posterior intervals.

信用区間の設定については、Chen & Shao (1999) およびIrony (1992) 参照。その他のベイズ統計の信用区間としては、最高事後密度 (HPD) 信用区間 (Lee (1997)) および中心事後密度信用区間がある。

6.4 Predictive probabilities

6.4 的中率

You may use predictive probabilities, a special type of posterior probabilities, in a number of ways:

的中率は特殊な事後確率であり、以下のような様々な方法で用いることができる。

Deciding when to stop a trial

試験中止時期の決定

If it is part of the clinical trial plan, you may use a predictive probability at an interim point as the rule for stopping your trial. If the predictive probability that the trial will be successful is sufficiently high (based on results thus far), you may be able to stop the trial and declare success. If the predictive probability that the trial

will be successful is small, you may stop the trial for futility and cut losses.

試験中止時期の決定が臨床試験計画の一部である場合、中間時点の的中率を試験中止基準として用いることができる。試験が成功するという的中率が十分に高い（それまでの結果に基づく）場合、当該試験を中止し、試験成功とすることができる。試験が成功するという的中率が低い場合は、当該試験は無益であるため中止し、損失を削減することができる。

Exchangeability is a key issue here: these predictions are reasonable only if you can assume the patients who have not been observed are exchangeable with the patients who have. This assumption is difficult to formally evaluate but may be more credible in some instances (e.g., administrative censoring) than others (e.g., high patient drop-out).

ここでは交換可能性が重要なポイントとなる。こうした推定は観察されていない患者が観察されている患者と交換可能であると仮定できる場合に妥当である。この仮定は、正式に評価することは難しいが、一部の例（管理打ち切りなど）においてはその他の例（高い患者脱落率）より信用できるかもしれない。

Predicting outcomes for future patients

今後の患者に関する転帰の推定

You may also calculate the predictive probability of the outcome of a future patient, given the observed outcomes of the patients in a clinical trial, provided the current patient is exchangeable with the patients in the trial. In fact, that probability answers the following questions:

- Given the results of the clinical trial, what is the probability that a new patient receiving the experimental treatment will be successful?
- What would that probability be if the patient were treated in the control group?

ある臨床試験の被験者と置換可能な患者が現れた場合、その臨床試験で観察された患者の転帰をもとに今後の患者の転帰の的中率を算出することもできる。実際に、そのような確率によって以下の点に対する回答を示すことができる。

- 臨床試験の結果に基づいた、新たな患者の治療が成功する確率
- 患者を対照群において治療した場合に治療が成功する確率

After device approval, these probabilities could be very useful in helping physicians and patients make decisions regarding treatment options.

機器の承認後、こうした確率は医師および患者が治療方法について決定する際に非常に有用となる。

Predicting (imputing) missing data

欠測値の推定（補填）

You may use predictive probabilities to predict (or impute) missing data, and trial results can be adjusted accordingly. There are also frequentist methods for missing data imputation.

的中率を用いて欠測値の推定（または補填）を行うことができ、それに従って試験結果を補正することができる。欠測値の補填には頻度論的手法を用いることもできる。

Regardless of the method, the adjustment depends on the assumption that patients with missing outcomes follow the same statistical model as patients with observed outcomes. This means the missing patients are exchangeable with the non-missing patients, or that data are missing at random. If this assumption is questionable, FDA recommends you conduct a sensitivity analysis using the prediction model. For examples of missing data adjustments and sensitivity analysis, see the Summary and Safety Effectiveness for PMA P980048, BAK/Cervical Interbody Fusion System, by Sulzer Spine-Tech.⁷

用いる手法にかかわらず、補正は転帰不明の患者が転帰の観察された患者と同じ統計学的モデルに従うという仮定に左右される。これは、転帰不明の患者は転帰の観察された患者と交換可能であること、あるいは観察データが無作為に欠損していることを意味する。この仮定に問題がある場合、FDAでは推定モデルを用いた感度解析の実施を推奨している。欠測値の補正および感度解析の例については、PMA P980048 Sulzer Spine-Tech製BAK/Cervical Interbody Fusion Systemの安全性および有効性に関する概要（Summary and Safety Effectiveness for PMA P980048, BAK/Cervical Interbody Fusion System by Sulzer Spine-Tech）を参照のこと。⁷

⁷ www.fda.gov/cdrh/pdf/p980048b.pdf

⁷ www.fda.gov/cdrh/pdf/p980048b.pdf

Predicting a clinical outcome from a surrogate

代用 (surrogate) からの臨床転帰の推定

If patients have two different measurements at earlier and later follow-up visits, you may make predictions for the later follow-up visit (even before the follow-up time has elapsed). Basing predictions on measures at the earlier visit requires that:

- some patients have results from both follow-up visits
- there is sufficiently high correlation between the early and the later measurement.

患者に前期と後期のフォローアップ来院時で2つの異なる値が測定された場合、(フォローアップ期間を過ぎる前でも) 後期フォローアップ来院について推定することができる。前期フォローアップ来院時の測定値に基づく推定には以下の条件が必要とされる。

- 一部の患者について前期、後期の両フォローアップ来院時の結果があること
- 前期および後期測定値の間に十分高い相関性があること

In this example, the outcome at the first time point is being used as a surrogate for the outcome at the second. This type of prediction was used to justify stopping the clinical trial of the INTERFIX Intervertebral Body Fusion Device.⁸

この例では、最初の時点の転帰が2度目の時点の転帰の代用として用いられている。この種の推定はINTERFIX Intervertebral Body Fusion Deviceの臨床試験中止の根拠を示すために用いられた。⁸

The surrogate may also be a different outcome; for example, for breast implants, rupture may be predictive of an adverse health outcome later.

代用として用いられるのが別の転帰の場合もある。例えば豊胸手術の場合、破裂から健康上の予後不良が推定できる。

Surrogate endpoints for predictive distributions should be validated. However, validation of a surrogate endpoint is a complex scientific and statistical issue that is outside the scope of this document.

予測分布の代用エンドポイントについてはその妥当性を検証すべきである。しかし、代用エンドポイントの妥当性検証は、本書の範囲外の複雑な科学的・統計学的問題である。

⁸ See Summary of Safety and Effectiveness for PMA P970015 at www.fda.gov/cdrh/pdf/p970015b.pdf.

⁸ See PMA P970015の安全性および有効性に関する概要 (www.fda.gov/cdrh/pdf/p970015b.pdf.) 参照

Model checking

モデル検証

FDA recommends you verify all assumptions important to your analysis. For example, an analysis of a contraceptive device might assume the monthly pregnancy rate is constant across the first year of use. To assess this assumption, the observed month-specific rates may be compared to their predictive distribution. You may summarize this comparison using a Bayesian p-value. For more information, refer to Gelman et al. (1996; 2004).

FDAでは、解析に重要なすべての仮定を実証することを推奨している。例えば避妊器具の解析では、月間妊娠率は器具使用の最初の1年を通して一定であると仮定される。この仮定を評価するには、観察された各月の妊娠率をその予測分布と比較する。この比較結果はベイズ法のp値を用いてまとめることができる。詳細についてはGelman et al. (1996、2004)を参照のこと。

You may also assess model checking and fit by Bayesian deviance measures as described in Spiegelhalter et al. (2002). Any predictive analysis assumes that patients for whom outcomes are being predicted are exchangeable with patients on whom the prediction is based. This assumption may not be valid in some cases. For example, patients enrolled later in a trial may have a different success rate with the device than those enrolled earlier if:

- ・ physicians have to overcome a learning curve in using the device
- ・ physicians form an opinion on whom the device treats best, and then enroll more or less favorable patients in the trial
- ・ an effective adjunctive therapy becomes available to the patients in the trial
- ・ an alternative treatment becomes available during the course of the trial, altering the characteristics of the patients who choose to enroll in the trial.

Spiegelhalter et al. (2002) が述べているように、ベイズ法の逸脱度測定によってモデル検証および適合度の評価を行うこともできる。どのような予測解析でも、転帰を推定される患者はその推定の根拠となっている患者と交換可能であると仮定される。この仮定は有効ではない場合もある。例えば、次のような場合、試験後期に登録された患者の機器使用成功率は、試験前期に登録された患者とは異なる可能性がある。

- 医師が当該機器の使用方法を習得する必要がある場合
- どの患者が当該機器による治療に適しているかを医師が評価し、おおよそ望ましい患者を試験に登録する場合

- 被験者が受けることのできる効果的な補助療法がある場合
- 試験実施中に代替治療の選択が可能となり、試験登録を選択した患者の特性が変わる場合

6.5 Interim analyses

6.5 中間解析

There is more than one method for analyzing interim results in a Bayesian trial. FDA recommends you specify the method in the trial design and ensure FDA agrees in advance of the trial. FDA may ask you to calculate the probability of a type I error through simulations before accepting a method. Although this is a frequentist calculation, it can help in evaluating the application of a method to a trial.

ベイズ法を用いた臨床試験の結果の中間解析には複数の方法がある。FDAでは、試験デザインでその方法を明示し、試験前にFDAの合意を得ることを推奨している。FDAでは、当該解析法の承認前にシミュレーションによって第1種の過誤の確率を算出するよう求めることがある。これは頻度論的な計算であるが、試験への解析方法の適用を評価するのに役立つと考えられる。

The following describes three specific Bayesian interim analysis methods:

ベイズ法による3つの具体的な中間解析法を以下に示す。

Applying posterior probability

事後確率の適用

One method stops the trial early if the posterior probability of a hypothesis at the interim look is large enough. In other words, the same Bayesian hypothesis test is repeated during the course of the trial.

中間解析時のある仮説の事後確率が十分に高い場合、1つの方法で試験を早期に中止する。すなわち、試験実施中に同じベイズ仮説検定を繰り返す。

Applying predictive distribution

予測分布の適用

Another method calculates at interim stages the probability that the hypothesis test will be successful. This method uses the Bayesian predictive distribution for patients yet to be measured. If the predictive probability of success is sufficiently high, the trial may stop early. If the predictive probability is very low, the trial may stop early for futility. This method was used in the submission of the INTERFIX Intervertebral Body Fusion Device.⁹

仮説検定が成功する確率を別の方法で中間解析時に計算する。この方法ではまだ測定されていない患者についてベイズ予測分布を用いる。成功の的中率が十分に高い場合には、試験は早期に中止できる。的中率が非常に低い場合は、試験は無益であるため早期に中止できる。この方法はINTERFIX Intervertebral Body Fusion Deviceの申請時に用いられた。⁹

Applying formal decision analysis

正式な意思決定分析の適用

A decision analysis method considers the cost of decision errors and experimentation in deciding whether to stop early. Carlin, Kadane, & Gelfand (1998) propose a method to approximate a decision analysis approach.

意思決定分析法では、試験を早期に中止するかどうかの決定において、意思決定過誤および実験にかかる費用を考慮する。Carlin, Kadane, & Gelfand (1998) は意思決定分析法に似た方法を提案している。

⁹ See the Summary of Safety and Effectiveness for PMA P970015 at <http://www.fda.gov/cdrh/pdf/p970015b.pdf>.

⁹ PMA P970015 安全性および有効性に関する概要 (<http://www.fda.gov/cdrh/pdf/p970015b.pdf>) 参照