
製薬業界向けガイダンス

制がん剤と生物製剤承認のための
臨床試験エンドポイント

米国保健社会福祉省
食品医薬品局
医薬品評価センター(CDER)
生物製剤評価研究センター(CBER)

2007年5月

臨床/医学

特定非営利活動法人日本臨床腫瘍学会 監訳

製薬業界向けガイダンス

抗がん剤と生物製剤承認のための 臨床試験エンドポイント

複本は下記で入手可能：

*Office of Training and Communications
Division of Drug Information, HFD-240
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857
(Tel) 301-827-4573
<http://www.fda.gov/cder/guidance/index.htm>*

または

*Office of Communication, Training, and
Manufacturers Assistance, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
(Tel) 800-835-4709 or 301-827-1800
<http://www.fda.gov/cber/guidelines.htm>*

米国保健社会福祉省
食品医薬品局
医薬品評価研究センター(CDER)
生物製剤評価研究センター(CBER)

2007年5月
臨床/医学

目次

I. 序文

II. 背景

- A. 有効性のための規制要件
- B. 腫瘍領域における過去の承認を裏付けるエンドポイント

III. 一般的なエンドポイントの考察

- A. 全生存期間
- B. 腫瘍評価に基づくエンドポイント
 - 1. 無病生存期間
 - 2. 客観的奏効率
 - 3. 無増悪期間および無増悪生存期間
 - a. TTP vs. PFS
 - b. 薬剤承認を支持するためのエンドポイントとしての PFS
 - c. PFS 試験デザインの問題点
 - d. PFS の解析
 - 4. 治療成功期間
- C. 症状評価を含むエンドポイント
 - 1. 個別的な症状エンドポイント
 - 2. 症状のデータが直面する問題
- D. バイオマーカー

IV. 臨床試験試験デザインの考察

- A. 単群試験
- B. 非劣性を証明するようにデザインされた試験
- C. 放射線保護剤と化学療法保護剤のための試験デザイン

V. 結論

付録 1：腫瘍測定データの収集

付録 2：PFS 解析において考慮すべき問題点

付録 3：PFS 解析のための表の例

付録 4：第三者による腫瘍エンドポイントのレビュー

製薬業界向けガイダンス¹ 制がん剤と生物製剤承認のための 臨床試験エンドポイント

本ガイダンスはこの問題に関する米国食品医薬品局(FDA)の現時点の考え方を示すものです。これにより何らかの権利を創出したり何者かに対して新たに権利を与えたりするものではなく、また、これはFDAや米国民を拘束するものでもありません。適用法令の規定する要件を満たすアプローチであれば、他のアプローチを用いても差し支えありません。他のアプローチについて検討されたい場合は、本ガイダンスの施行に責任を有するFDAのスタッフにご連絡ください。FDAのどのスタッフに連絡を取ればよいか不明な場合は、本ガイダンスの表紙に書かれている電話番号にご一報ください。

I. 序文

本文書は、計画的ながんエンドポイントガイダンスシリーズの最初のガイダンスである。本ガイダンスは、新薬承認申請(NDA)、生物製剤承認申請(BLA)、または追加申請における有効性の申し立てを裏付けるため食品医薬品局(FDA)に提出されるがん臨床試験のエンドポイントについて、申請者に提言を行なう。²本ガイダンスは背景情報も提供し、一般的な規制原理について論じる。本ガイダンスにおいて論じるエンドポイントは、既存のがん患者を治療するための薬剤のためのものである。本ガイダンスは、がんの新たな発生を予防または低減する薬剤のためのエンドポイントについて言及するものではない。

FDAは、「抗癌剤諮問委員会(ODAC)」の前の公開ワークショップおよび話し合いを含む過程を通じ、腫瘍エンドポイントについてのガイダンスを作成中である。³今後の各ガイダンスでは、個別のがん(例、肺癌、結腸がん)のためのエンドポイントに焦点を絞り、医薬品承認または表示申し立ての内容を裏付ける予定である。

本ガイダンスを含むFDAのガイダンス文書は、法的に強制しうる責任を定めるものではない。むしろ、ガイダンスはある問題に関してFDAの現時点での考え方を示すものであり、具体的な規制または法的要件が挙げられていない限り、単に提言とみなされる。FDAのガイダンスにおける *should* という単語の使用は、ある事柄を提案すること(suggest)、または勧告すること(recommend)を意味し、要求すること(require)を意味するものではない。

¹ 本ガイダンスは、食品医薬品局の生物製剤評価研究センター(CBER)との協力で、薬剤評価研究センター(CDER)の抗がん剤担当部署(Division of Drug Oncology Products)と生物学的抗がん剤担当部署(Division of biologic Oncology Products)により作成された。

² 本ガイダンスの目的に沿い、特記されない限り薬剤(drugs)という場合はすべてヒト用医薬品と生物学的製剤の両者を含む。

³ 本文書は下記のウェブページで閲覧できる

http://www.fda.gov/cder/drug/cancer_endpoint/default.htm

II. 背景

本文書は法的拘束力を持たない勧告を含む

臨床試験のエンドポイントは様々な目的を果たす。従来の抗がん剤開発において、早期臨床試験は、安全性を評価し腫瘍縮小などの薬剤の生物学的活性を示すエビデンスを明らかにする。後期の有効性評価試験のエンドポイントは通常、ある薬剤が生存期間延長または症状改善などの臨床的有益性を提供するかを評価する。以下の節においては、有効性のための一般的規制要件と、それが抗がん剤承認のためのエンドポイント選択にどのように影響しているかについて論じる。後節では、これらのエンドポイントについて詳しく述べ、それが疾患活動性または種々の臨床状況における臨床的有益性の尺度となるかどうかについて論じる。

A. 有効性のための規制要件

新薬は有効性を示さなければならないという要件は、連邦食品医薬品化粧品法(the Federal Food, Drug, and Cosmetic Act)の1962年改正に基づく。この法律は有効性についての実質的エビデンスを要求しており、このエビデンスは適切な比較対照試験に由来するものでなければならないと明記している。同様に、公衆衛生法(the Public Health Service Act)は、生物製剤は安全で、不純物を含まず、効力を有して(potent)いなければならないと要求している。医薬品承認を裏付ける臨床的有用性は重要な臨床的アウトカム(例、生存期間延長、症状改善)を含むが、確立された代替エンドポイント(例、血圧、血中コレステロール)に対する作用も含む。

1992年に交付された迅速承認制度(21 CFR Part 314, Subpart H と 21 CFR Part 601, Subpart E)により、重篤または生命にかかわる疾患治療を意図し、現在行なわれている治療法を上回る改善を示す、またはこれまで治療法が存在しない場合の薬剤または生物製剤の承認のため、追加エンドポイントの使用が可能になった。この状況において、FDAは、(疫学的、治療的、病態生理学的、またはその他のエビデンスに基づき)臨床的有用性を適切な尤度で予測しうると考えられる代替エンドポイントの有効性に基づき、承認することができる。そのような代替は、心血管疾患のための血圧またはコレステロールなど恒常的に使用されている代替ほど十分には確立していない。薬剤は、製造者が実際の臨床的有用性を証明し説明するための臨床試験を実施する場合には、迅速承認制度の下で承認される。市販後臨床試験により臨床的有用性を証明できない場合、または申請者が、要求される試験実施において正当な努力を怠った場合には、当該薬剤は迅速に市場から回収される。以下の考察において、通常承認(regular approval)という用語は臨床的有用性の証明に基づく薬剤承認の長期ルートを表す。この用語は、臨床的有用性を適切に予測すると考えられる代替エンドポイントの使用を伴う迅速承認(accelerated approval)と区別される。

臨床試験の望ましい件数を含め、薬剤承認を裏付けるために不可欠なエビデンスについては、業界向けガイダンス *FDA Approval of New Cancer treatment Uses for Marketed Drug and Biological Products*⁴ および 1997年のFDA近代化法⁵において論じられている。

⁴ ガイダンスは定期的に更新されている。最新版ガイダンスであることを確認するため、CDERガイダンスウェブページをチェックされたい：<http://www.fda.gov/cder/guidance/index.htm>

⁵ <http://www.fda.gov/cder/fdama/default.htm>

本文書は法的拘束力を持たない勧告を含む

ほとんどの場合、FDA は、適切な比較対照試験を少なくとも 2 件実施するよう勧告している。ある種の場合には、1 試験のエビデンスでも十分とみなされる（例、1 回の多施設共同試験で生存に対する効果などの重要な臨床的有用性について高度に信頼でき統計的に強力なエビデンスが得られる場合、および 2 回目の試験で結果を確認することが現実的または倫理的に不可能である場合）。特殊な悪性腫瘍の特定の病期である患者の治療のため承認される薬剤の場合、1 試験のエビデンスは、同一がんの異なる病期の治療にとって efficacy supplement（有効性補完）を裏付けるために十分とみなされる。

B. 腫瘍領域における過去の承認を裏付けるエンドポイント

通常承認の場合、申請者は、臨床的有用性の直接的エビデンスまたは臨床的有用性の確立した代替エンドポイントにおける改善を示すことが不可欠である。腫瘍学においては、生存期間の改善は臨床的有用性の適切な尺度とみなされている。さらに、スポンサーは抗がん剤承認のための他のエンドポイントを使用している。1970 年代、FDA は通常、X 線検査または理学的検査による腫瘍評価により確認される客観的奏効率(ORR)に基づいて抗がん剤を承認した。1980 年代早期、ODAC との話し合いの後に FDA は、生存期間改善、患者の QOL 改善、身体機能改善、または腫瘍関連症状の改善など、より直接的な臨床的有用性のエビデンスに基づいて抗がん剤の承認を行なうと決定した。これらの有用性は必ずしも常に ORR によって予測されるものではなく、また ORR と関連するものでもない。

以後の 10 年間、臨床的有用性の確立した代用としていくつかのエンドポイントが使用された。無病生存期間(DFS)改善は、再発時に大部分の患者ががんの症状を呈すると予測される特定のアドジュバント状況における薬剤承認を支持した。白血病では、感染症、出血、血液製剤によるサポートが少ない持続的な完全奏効は、臨床的有用性の確立したエンドポイントとみなされた。十分立証された高い ORR は特定の固形腫瘍における通常承認を支持することができるが、承認の決定を下す場合には、奏効期間、腫瘍関連症状の緩和、薬剤毒性も考慮する(O'Shaughnessy and Wittes et al., 1991, Commentary Concerning Demonstration of Safety and Efficacy of Investigational Anticancer Agents in Clinical Trials, J Clin Oncol, 9:2225-2232)。例えば、乳がんのためのホルモン製剤の無作為化試験は、通常承認を裏付けるエンドポイントとして ORR を使用している。ORR 改善および奏効期間が十分であることと併せての腫瘍関連症状改善は、いくつかの臨床状況において通常承認を支持している。

迅速承認の代替エンドポイントは、臨床的有用性が適切な尤度で予測されなければならない(21 CFR Part 314, Subpart H と 21 CFR Part 601, Subpart E)。そのような薬剤は、既存の治療法を上回る有用性も提供しなければならない(21 CFR Part 314, Subpart H と 21 CFR Part 601, Subpart E)。⁶ORR は迅速承認を裏付けるために最も一般的に使用される代替エンドポイントである。腫瘍縮小効果はがん治療の指針において腫瘍医に広く受け入れられている。ORR は薬剤の効果に直接的に帰せられるため、治療法が存在しない難治性腫瘍患者において施行される単群試験は ORR についての正確な評価を提供する。

6 業界向けガイダンス「利用できる治療法」(<http://www.fda.gov/cder/guidance/index.htm>) 参照。

III. 一般的なエンドポイントの考察

本節では、抗がん剤開発における一般的問題についての概観を述べる。一般的に使用されるがんのエンドポイントに関する考察の後に、これらのエンドポイントを使用するがん臨床試験デザインに関する問題の考察を行なう（今後のガイダンスでは個別の治療適応症エンドポイントについて論じる予定である）。本節で論じるエンドポイントは、全生存期間、腫瘍評価（例、DFS、ORR、完全奏効、無増悪期間 [TTP]、無増悪生存期間 [PFS]）に基づくエンドポイント、および症状評価に基づくエンドポイントである。表 1 は抗がん剤承認におけるエンドポイントの比較である。有効性エンドポイントの適正な分析に関連する多くの問題点は、業界向け ICH ガイダンス *E9 Statistical Principles for Clinical Trials* に取り上げられている。

表 1. 抗がん剤承認の重要なエンドポイントの比較

エンドポイント	規制承認における位置づけ	試験デザイン	長所	短所
全生存期間	通常承認のための臨床的有用性	<ul style="list-style-type: none"> 無作為化試験必須 盲検は必須ではない 	<ul style="list-style-type: none"> 有用性について普遍的に受容される直接的尺度 容易に測定される 正確に測定される 	<ul style="list-style-type: none"> 大規模試験を含む可能性 クロスオーバーまたは順次型投与 (sequential therapy) により影響されうる 非がん死を含む
症状エンドポイント (患者が申告するアウトカム)	通常承認のための臨床的有用性	<ul style="list-style-type: none"> 無作為化盲検試験 	<ul style="list-style-type: none"> 直接的な臨床的有用性についての患者の視点 	<ul style="list-style-type: none"> 盲検は困難な場合が多い データは欠落または不完全な場合が多い 小さい変化の臨床的有意性は不明 複合分析 妥当性の証明された測定手段がない
無病生存期間	迅速承認または通常承認のための代替*	<ul style="list-style-type: none"> 無作為化試験必須 盲検が望ましい 盲検レビューが勧告される 	<ul style="list-style-type: none"> 生存試験と比較して小さいサンプルサイズと短期のフォローアップが必要 	<ul style="list-style-type: none"> すべての状況における生存の代替としての統計的な検証がされていない 正確に測定されない；評価バイアスの傾向、特にオープン・ラベル試験で 試験によって定義が異なる

本文書は法的拘束力を持たない勧告を含む

表 1、続き

客観的奏効率	迅速承認または通常承認のための代替*	<ul style="list-style-type: none"> ・単群または無作為化試験を使用 ・比較試験では盲検が望ましい ・盲検レビューが勧告される 	<ul style="list-style-type: none"> ・単群試験で評価 ・生存試験と比較して早期および小規模試験で評価される ・効果は自然歴ではなく薬剤に帰せられる 	<ul style="list-style-type: none"> ・有用性の直接的尺度ではない ・薬剤活性の包括的尺度ではない ・効果があるのは患者のごく一部のみ
完全奏効	迅速承認または通常承認のための代替*	<ul style="list-style-type: none"> ・単群または無作為化試験を使用 ・比較試験では盲検が望ましい ・盲検レビューが勧告される 	<ul style="list-style-type: none"> ・単群試験で評価 ・持続的な完全奏効は臨床的有用性を表す ・生存試験と比較して早期および小規模試験で評価される 	<ul style="list-style-type: none"> ・すべての症例における有用性の直接的尺度という訳ではない ・薬剤活性の包括的尺度ではない ・効果があるのは患者の一部
無増悪生存期間（全死亡を含む）または無増悪期間（増悪前の死亡打ち切り）	迅速承認または通常承認のための代替*	<ul style="list-style-type: none"> ・無作為化試験必須 ・盲検が望ましい ・盲検レビューが勧告される 	<ul style="list-style-type: none"> ・生存試験と比較して小さいサンプルサイズと短期のフォローアップが必要 ・安定疾患の測定を含む ・クロスオーバーまたは後続治療に影響されない ・一般的に、客観的および定量的評価に基づく 	<ul style="list-style-type: none"> ・すべての状況における生存の代替としての統計的検証がされていない ・正確に測定されない；評価バイアスの傾向、特にオープン・ラベル試験で ・試験によって定義が異なる ・頻繁な X 線などによる評価 ・治療群間で平衡したタイミングで評価する

*迅速承認または通常承認のための代替エンドポイントとしての適性は、エフェクトサイズ、効果持続期間、およびその他利用できる治療法の有益性など、他の因子に高度に依存する。

A. 全生存期間

全生存期間とは、無作為化から全原因による死亡までの時間と定義され、**intent-to-treat** の原則に沿った全患者において測定される。生存期間は最も信頼できるがんのエンドポイントとみなされ、生存期間について十分評価できる試験が実施される場合、生存期間は通常望ましいエンドポイントである。このエンドポイントは正確で測定しやすく、死亡日によって証明される。エンドポイント測定にバイアスは生じない。生存期間改善は臨床的有用性を評価するためのリスク便益分析として解析すべきである。

本文書は法的拘束力を持たない勧告を含む

全生存期間は無作為化比較試験で評価すべきである。過去の試験によるデータは、時間依存性エンドポイント（例、全生存期間、PFS）にとって信頼性が高いことは稀である。歴史対照群と現在の治療群におけるアウトカムの明白な相違は、患者選択、画像撮影技術の改良または支持療法の改善を含む、薬剤治療以外の差によって生じる。無作為化試験は、直接的にアウトカムを比較することによりこれらの相違の影響を最小化する。全生存期間における統計的に有意な改善の立証は、毒性像が受容できるものであれば統計的に有意とみなされ、新薬承認を支持することが多い。

生存試験を施行、解析する際の困難は、大規模試験においてフォローアップ期間が長期であること、後続のがん治療によって生存解析が交絡を受ける可能性があることなどである。

B. 腫瘍評価に基づくエンドポイント

本節では、腫瘍評価に基づくいくつかのエンドポイントについて論じる。これらのエンドポイントには、DFS、ORR、TTP、PFS、および治療成功期間(TTF)が含まれる。これらの時間依存性エンドポイントに関するデータの収集と解析は、間接的評価、計算、および推定値（例、腫瘍の計測値）に基づく。PFS データ収集と解析は、付録 3 に示す表で補足する。

腫瘍評価エンドポイントの選択には 2 つの判断が含まれる。まず、そのエンドポイントが、迅速承認と通常承認のいずれかを裏付けるのかどうかを明らかにすべきである。次に、そのエンドポイントが腫瘍エンドポイント評価においてバイアスをもたらす可能性、または不確実性について評価すべきである。唯一の有効性のエビデンスとして腫瘍の計測に基づくエンドポイントに重きを置く試験を用いる薬剤申請は、二次試行による補強証拠が必要である。腫瘍計測の精度は腫瘍の状況によって異なる可能性がある。奏効率の判定に使用する腫瘍計測値は、境界線のない場合（例、悪性中皮腫、膵がん、脳腫瘍）では不正確になりうる。

試験の主要エンドポイントが腫瘍計測値（例、PFS または ORR）に基づいている場合、腫瘍エンドポイント評価は通常、試験治療について知らされていない中央判定医(central reviewer)が検証すべきである（付録 4 参照）。この尺度は、試験自体が盲検でない場合には特に重要である。FDA は、中央判定の過程を検証するためスキャンのサンプルを調査してもよいと思われる。データ収集に関するその他の詳細は付録 1 に記載する。腫瘍エンドポイント評価（特に PFS または DFS の場合）の中央による独自の検証は、無作為化試験が盲検である場合（有害事象プロファイルによって施行中の試験が実質的に非盲検化されるのでない限り）、または感受性解析で観測者によるバイアス（特に DFS の場合）がないことを裏付ける大規模無作為化試験でエフェクトサイズが大きい場合には、必ずしも必要でない。

1. 無病生存期間(DFS)

一般的に、DFS は無作為化時点から腫瘍再発または全原因による死亡までの期間と定義される。このエンドポイントを最も頻繁に使用するのは、根治手術または放射線療法後のアジュバント療法においてである。DFS は、大半の患者が化学療法で完全奏効に達する場合の重要なエンドポイントでもある。全生存期間はほとんどのアジュバント療法における従来のエンドポイントであるが、DFS は生存期間が延長される状況における重要なエンドポイントであり、生存期間のエンドポイントは非実

本文書は法的拘束力を持たない勧告を含む

用的なものになる。乳がんの術後補助ホルモン療法、結腸癌の術後補助療法、および乳がんの術後補助化学療法の承認においては、DFS が基本として用いられてきた。標準的な抗がん剤療法と比較するとホルモン療法の副作用は最小限であり、したがってリスクと便益の関係が好ましいものとなる。DFS は臨床的有用性の代替となり、また臨床的有用性の直接的エビデンスを提供する。この判定は、効果の規模、そのリスクと便益の関係、および疾患の状況に基づいている。しかし、延命効果がすでに確立している疾患状況においては、DFS が臨床的に有用とみなされるとは考えにくい。2003年12月のODACのコンセンサスでは、この便益の規模がアジュバント治療の毒性に勝るものである場合には、DFS 延長は臨床的有用性を示すとされた。2004年5月、ODACは、DFS を外科アジュバント療法における結腸がん治療薬にとって受容できるエンドポイントとみなすように勧告した。⁷

DFS を潜在的エンドポイントとして評価する際に考慮すべき重要な事項として、推定される治療効果のサイズと標準療法において証明されている便益がある。プロトコルではDFS の定義とフォローアップ検査および通院予定を慎重に定めておくべきである。様々な理由で予定外の評価が生じることもあり、予定外評価の回数、タイミング、理由における試験群間の差はバイアスをもたらすことがある。バイアスは、治療群の割付について患者と治験責任医師を盲検化することにより最小化できる。予定外の評価によるバイアスの内在する影響は、事象の発生時期とは無関係にフォローアップ期間全体にわたる事象の総数を比較解析することによって評価できる。

DFS の定義は、特に先行の腫瘍増悪が文書化されずに死亡が記録された場合に複雑化する。これらの事象は疾患再発または打ち切り事象と記録される。死亡についての統計解析の手法はすべて何らかの制約を有しているが、すべての死亡（全原因による死亡）を再発とみなせばバイアスは最小化できる。DFS は、特に観察なしで長期間経過した後に死亡した患者の場合には、この定義によると過大評価される。長期間のフォローアップ来院頻度が試験群間で異なる場合、または脱落例が毒性のために任意ではない場合に、バイアスが生じうる。一部の解析はがん関連死をDFS 事象と計数し、非がん死の追跡を打ち切っている。この手法は死因の帰属にバイアスをもたらす。さらに、死亡時または最終来院時に患者の追跡を打ち切る手法はすべて、打ち切り患者と非打ち切り患者の再発リスクは同じと仮定している。

2. 客観的奏効率

ORR は、前もって定めた量の腫瘍サイズ縮小を最短期間示した患者の割合と定義される。奏効の持続期間は通常初回奏効時点から文書化された腫瘍増悪までとされる。一般に、FDA はORR を部分奏効と完全奏効の合計と定義してきた。このように定義すると、ORR は薬剤の抗腫瘍活性の直接的尺度であり、単群試験において評価可能である。不変はORR の成分とみなすべきではない。不変は疾患の自然歴を反映する一方、腫瘍縮小は治療の直接的効果である。また、不変はTTPまたはPFD解析（下記参照）によってより正確に評価される。できれば、標準化された基準を用い

⁷ 写しは下記のウェブページで閲覧できる

http://www.fda.gov/cder/drug/cancer_endpoint/default.htm

本文書は法的拘束力を持たない勧告を含む

て奏効を確認するとよい。様々な奏効基準が適切とみなされている(例、RECIST 基準)(Therasse and Arbuck et al., 2000, New Guidelines to Evaluate Response to Treatment in Solid Tumors, J Natl Cancer Inst, 92:205-16)。奏効基準は試験開始前にプロトコールに定めておく必要がある。ORR の有意性はその規模と持続期間、および完全奏効(腫瘍のエビデンスが検知されない)の割合によって評価する。

3. 無増悪期間および無増悪生存期間

TTP と PFS は薬剤承認のための主要エンドポイントの役割を果たしている。TTP は無作為化時点から客観的な腫瘍増悪までの時間と定義される；TTP は死亡を含まない。PFS は、無作為化から客観的な腫瘍増悪または死亡までの時間と定義される。腫瘍増悪の正確な定義は重要で、プロトコールに慎重に述べておく必要がある。

a. TTP vs. PFS

TTP と比較して PFS は規制エンドポイントとして望ましい。PFS は死亡を含んでいるため、全生存期間とより良く相関する。TTP 解析では、死亡時点か、または情報打ち切り(試験からの消失の非任意パターン)を示すそれ以前の診察時点において、死亡により追跡が打ち切られる。PFS は、患者の死亡は腫瘍増悪と任意に関連するとみなす。しかし、死亡例の大半ががんと無関係である場合には、TTP は受容できるエンドポイントである。

b. 薬剤承認を支持するためのエンドポイントとしての PFS

表 1 にエンドポイントとして PFS を使用する際の長所と短所を述べる。PFS は腫瘍増殖を反映することができ、生存利益の判定前に評価できる。その判定は後続治療によって交絡をきたすことがない。任意のサンプルサイズについて、PFS に対する作用の規模は全生存期間に対する作用より大きくなりうる。しかし、多数の各種既存悪性腫瘍にとって生存期間の代用としての PFS の正式な有効性の立証は困難である。データは通常、生存期間に対する作用と PFS の相関性の確実な評価を可能にするには不十分である。がん試験は小規模な場合が多く、既存薬に証明されている延命効果は通常ささやかなものである。承認を与えるためのエンドポイントとしての PFS の役割は種々のがんの状況によって異なる。PFS 改善が直接的な臨床的有益性を示すのか、それとも臨床的有益性の代用であるのかは、効果の規模と、既存の治療法と比較した新しい治療法のリスク便益の考察に依存する。

c. PFS 試験デザインの問題点

PFS の評価、測定および解析を行う方法論はプロトコールと統計解析計画(SAP)に詳細に述べておくべきである。腫瘍増悪基準をプロトコールに慎重に定義しておくことも重要である。増悪を定義する際の標準的な規制規準はない。申請者は、RECIST 基準を含む様々な基準を使用してきた。発表されたほとんどの PFS 基準に述べられている幅広いアウトラインは、プロトコールと SAP において詳細を補足すべきである。来院と放射線学的評価は、系統的バイアスを防止するため 2 試験群間で対称的に行われなければならない。可能であれば、試験は盲検にすべきである。盲検化は、患者または治験責任医師の評価が増悪エンドポイントの成分として加えられる場合には特に重要である。少なくとも、評価は、通常放射線医と臨床医から構成される盲検化された第三者判定チームに委ねる必要がある。FDA と申請者は下記の項目について予め合意しておかねばならない：

本文書は法的拘束力を持たない勧告を含む

- ・ 試験デザイン
- ・ 増悪の定義
- ・ 症例報告書(CRF)に記録するデータ
- ・ SAP
- ・ 欠測データを処理する方法論と打ち切りの手法
- ・ 該当する場合には、エンドポイントの独立レビュー委員会(IRC)の運営手続き (付録 4 参照)

d. PFS の解析

欠測データがあると PFS 解析は複雑になる。プロトコルでは各患者についての適正な評価のための来院の定義をしておかねばならない (すなわち、1 回の来院時に計画に基づく腫瘍評価すべてを行う)。解析計画では、各治療群におけるフォローアップの適正さの比較について概説しておくべきである。不完全および/または欠測しているフォローアップ来院および打ち切り手法を解析する方法論もプロトコルに明記しておく。解析計画では、結果の確実性を評価するため、主要解析と 1 回以上の感受性分析について明記しておくべきである。欠測データがある解析はすべて問題を含むものの、結果は、主要解析と感受性分析の両方に類似する結果がみられれば強化される。評価には、長期間にわたる追跡不能症例の死亡数を含める必要がある。そのような死亡数に不均衡があれば、フォローアップが短い治療群における PFS の過大評価によって PFS 測定値にバイアスが生じることがある。

増悪データは複数の情報源 (予定外来院時の理学的検査および各種 X 線スキャン) および様々な時点から収集できるため、評価のための来院時毎のデータ収集は、予定された来院日前後の特定の短期間に限定する必要がある。データを長期間にわたって収集する場合には、事象の日付および打ち切り日の判定に困難が生じることがある。我々は、それまで評価の欠落がなく、前回の放射線学的評価によって増悪していないことが確認された日に打ち切りが行なわれていない状況で増悪が観測された場合は、最も早い日を増悪日とするように勧告している。付録 3 に、主要解析および感受性分析に利用できるポテンシャル PFS 解析のための一連の表を収載する。PFS データ収集と解析のための計画は、第 2 相臨床試験終了時の会合で FDA と協議し、特別なプロトコル評価において確認しておくべきである。

4. 治療成功期間

TTF は、無作為化時点から、疾患増悪、治療毒性および死亡を含む全原因による治療中止までの複合エンドポイント測定時間と定義される。TTF は薬剤承認の規制エンドポイントとしては推奨されていない。TTF はこれらの補足変数と効果を明確に区別するものではない。規制エンドポイントは、薬剤の有効性と毒性、患者または医師による薬剤の使用中止、または患者の不耐性を明確に区別しなければならない。

C. 症状評価を含むエンドポイント

症状改善は臨床的有用性とみなされる。FDA 薬剤承認は患者の症状評価および／または症状改善を表す身体兆候（例、体重増、滲出液の減少）を主要有効性エンドポイントとして使用している。しかし、世界的に用いられている健康に関する QOL(HRQL)の尺度は、抗がん剤承認においては主要有効性エンドポイントとして利用されていない。HRQL の評価尺度とその妥当性は、業界向けガイダンス *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims* の草稿で論じられている。⁸兆候と症状の改善または QOL 評価を抗がん剤承認裏付けのための主要エンドポイントとして使用するには、FDA が腫瘍症状が改善したことで薬剤毒性がないことを識別できるべきである。世界的な HRQL 評価尺度に基づく明白な効果の優位性は、効果というより単に毒性がないことを示している可能性がある。

1. 個別的な症状エンドポイント

がん症状無増悪期間(*Time to progression of cancer symptoms*)は TTP と同様のエンドポイントであり、代替エンドポイントとなる可能性があるものというよりむしろ臨床的有用性の直接的尺度である。すでに論じたように、症状増悪までの期間を評価する際には増悪測定上の諸問題（例、評価欠落）も存在する。がん試験が盲検であることは少ないため、評価はバイアスを被ることが考えられる。腫瘍増悪からがん症状発現までの時間が遅延することもありうる。症状エンドポイントに達する前に代替療法を開始することも多く、この解析に交絡をもたらす。がん試験の多くはがん症状が最小限である患者で行なわれ、腫瘍症状と薬剤毒性を識別することが困難である。

複合症状エンドポイントは、臨床的重要性が類似した成分を含むべきであり、結果はひとつの成分だけに帰せられるべきではない。例えば、薬剤は、複合的な有用性エンドポイントに基づいて、骨格転移がん患者の治療のために承認されている。骨格関連事象は、病理骨折、骨への放射線療法、骨の手術、脊髄圧迫と定義されている。

症状緩和を立証するには適切な群の選択が不可欠である。試験のベースラインにおいて症候性である患者はカテゴリー別症状の反応解析により評価できる。ベースラインにおいて無症候である患者では、初発症状までの時間(*time-to-first-symptom*)解析を利用できる。患者が試験薬の使用を中止または新しい薬剤を使用し始めた場合には、初発症状の立証までフォローアップが継続されていれば、症状増悪はそのまま評価可能である。

2. 症状のデータが直面する問題

データの欠測および評価の回数が少ないことは、特にオープン・ラベル試験では症状データの評価を複雑にする。薬剤毒性または腫瘍増悪による治療中止は症状データ欠測の一因である。患者が治療を

⁸ 最終稿では、このガイダンスは、この問題についての FDA の最新の考え方を表すものになる。最新版ガイダンスの確認は、CDER ガイダンスウェブページをチェックされたい：

<http://www.fda.gov/cder/guidance/index.htm>

本文書は法的拘束力を持たない勧告を含む

中止しても、解析情報を提供するためにデータ収集用紙で情報収集を継続することが望ましい。複数の症状に関するデータ収集は多重性に関して予め対応し、SAPにおいて必要な統計的調整を明記しておくべきである。

D. バイオマーカー

一般的に、血液または体液で分析するバイオマーカーは抗がん剤承認の主要エンドポイントとしての役割は果たさないが、血液と尿で測定するパラプロテイン濃度は骨髄腫の奏効基準の一部として利用されている。既存の検査の妥当性を確立し、バイオマーカー値の改善が臨床的有用性を予測できるかを明らかにするためには、今後の研究が待たれる。

FDAは腫瘍マーカーを複合エンドポイントの成分として認めている。卵巣がん患者では、CA-125の著明な上昇を伴って、ある種の臨床事象（一般状態の著しい低下、または腸閉塞）が発現することは増悪とみなされる。また、バイオマーカーは、予後因子の特定と、試験デザインにおいて考慮すべき患者や層別因子の選択に有用である。

IV. 臨床試験デザインの考察

FDAは、21 CFR 314.126に従い、規定に述べられているとおり“適切な比較対照試験”に由来する有効性の実質的エビデンスに基づいて薬剤を承認する。試験は対照との有効な比較ができなければならない。有効性立証において最も信頼できる手法は、盲検無作為化比較試験において臨床的に意味のあるエンドポイントの統計的に有意な改善を示すことである。以下の節では、がん試験のデザインに関するいくつかの問題点について論じる。

A. 単群試験

治療法が存在せず、重大な腫瘍縮小が試験薬に帰せられると推測される場合に、FDAはときに、迅速承認を裏付ける実質的エビデンスとして、単群試験で観察されたORRと奏効期間を支持する。奏効率は、急性白血病などの通常承認を求める状況で利用され、完全奏効は輸血の必要性の低下や感染症の減少、生存期間の延長を伴う。単群試験は、生存期間やTTP、またはPFS等の事象発生までの時間のエンドポイントを十分説明するものではない。様々な種類のがんの自然歴はまさに多様であるため、事象までの時間エンドポイントを評価するには無作為化試験が必要である。

B. 非劣性を証明するようにデザインされた試験

非劣性(NI)試験は、新薬が、前もって規定した量（非劣性マージン）により、標準レジメン（実薬対照）と比較して有効性が劣らないことを示すことによって、その新薬の有効性を証明する(Temple and Ellenberg, 2000, Placebo-Controlled Trials and Active-Control Trials in the Evaluation of New Treatments, Part 1: Ethical and Scientific Issues, Ann Intern Med, 133(6):455-63)。⁹この非劣性マージンは、有効な実薬対照に比べて、臨床的に容認される範囲の有効性の低下にとどまるべきであ

本文書は法的拘束力を持たない勧告を含む

る。標準レジメンは特徴が十分解明された臨床的有用性（延命効果）を有するべきである。新薬が実薬対照と比較して非劣性マージン以上に劣る場合には、効果がないと推測される。

NI 試験は外部（過去の）データに依存して実薬対照による治療のエフェクトサイズを確立する。がん試験では、この効果（エフェクト）の特徴は十分解明されていない。NI 試験は恒常性についての仮定にも依存する。この仮定は、実薬対照の効果は過去の試験と現在の試験の間においても恒常的に持続しているという仮定である。これは、患者群の特徴や支持療法、および現試験と実薬対照の効果が明らかにされた過去のデータ間の評価方法も恒常的であると仮定している。実薬対照に推定される治療効果のサイズは、過去の試験の包括的なメタアナリシスに基づくべきである。これらの試験は、プラセボ群と比較した実薬対照の効果を確実に再現できると思われる。NI 試験実施における難点は、実薬対照の効果の推定と、維持される効果量（NI マージン）の確認である。NI 試験は通常、優越性試験と比較してサンプルサイズが大きく、過去の臨床試験と同じ効果を生むことが多い。さらに、後続療法と実薬対照群へのクロスオーバーによりすべての NI 解析に交絡が生じる。生存期間以外のエンドポイントを有する NI 試験には問題が多い。

C. 放射線保護剤と化学療法保護剤のための試験デザイン

放射線保護剤と化学療法保護剤は、治療による毒性軽減のためにデザインされた薬剤である。これらの試験には通常 2 つの目的がある。ひとつは、がん治療の毒性軽減の評価である。第二の目的は、抗がん活性が保護剤によって減弱するかどうかを明らかにすることである。第二の目的では通常、代替エンドポイント、例えば、全生存期間ではなく ORR または TTP を検討する。

V. 結論

本ガイダンスにおいて概説した一般原理は、申請者が市販承認申請のためのエンドポイントを選択する際に役立つと思われるが、我々は、申請者が NDA または MLA 申請のプロトコルを提出する前に FDA との会議をもつように勧告する。FDA は、この会議に腫瘍医、統計学者、臨床薬理学者、および多くの場合外部の専門コンサルタントという多分野からなる FDA チームが参加することを保証する。申請者はこの会議の後にプロトコルを提出し、エンドポイントと薬剤市販承認申請を裏付けるプロトコルデザインの受容性確認を提供する特別なプロトコル評価を要請することができる。¹⁰ 当然ながら、最終的には、市販承認は臨床試験のデザインだけではなく、薬剤市販承認申請における全試験の結果とデータの FDA によるレビューに基づいて行なう。

⁹ 業界向け ICH ガイダンス *E10 Choice of Control Group and Related Issues in Clinical Trials* 参照 (<http://www.fda.gov/cder/guidance/index.htm>)

¹⁰ 業界向けガイダンス *Special Protocol Assessment* 参照 (<http://www.fda.gov/cder/guidance/index.htm>)

本文書は法的拘束力を持たない勧告を含む

付録 1： 腫瘍測定データの収集¹¹

以下は、腫瘍測定データに関する重要な考慮すべき事項である。われわれは以下のとおり勧告する。

- CRF および電子データにより、治療前のベースライン来院時に特定された標的病変を記録すること。こうした病変をレトロスペクティブに特定することは、信頼性を欠く。
- 腫瘍病変に固有の識別文字または番号を付与すること。これにより、ひとつの解剖学的部位に発生する複数の腫瘍の鑑別ができ、ベースライン時に測定した腫瘍とフォローアップ時に測定した腫瘍のマッチングが可能になる。
- フォローアップの不可欠な時期における完全なデータ収集を確実にするメカニズムを整える。CRF では、標的病変のすべてがベースラインの評価を確実に受け、ベースラインとフォローアップ時に要求される全検査において同一の撮影または測定方法が使用されるように留意する。
- CRF には、来院のたびにスキャンが実施されたかどうか明記する欄を備えること。
- 病変が完全に消失した場合にはゼロと記録する。そうでなければ、欠測値と病変喪失を区別することができない。
- フォローアップ検査で初発部位と新たな発生部位の両者における新しい病変を、適宜検出する。新しい病変の発生と位置を CRF および申請する電子データに記録すべきである。

¹¹ 当付録の趣旨に従い、腫瘍データとは画像ではなく SAS トランスポートファイルのデータを指す。一般的に、画像は NDA または BLA では提出しないが、レビュー過程で FDA がレビューすることがある。

付録 2： PFS 解析において考察すべき問題点

プロトコールと SAP では PFS の主要解析について詳しく定めておくべきである。この解析には、エンドポイントについての詳細な説明、腫瘍評価に適切なモダリティ、IRC のための手続きなどバイアス最小化のための手順を含める。一つまたは二つの二次解析を指定し、試験実施において予想される問題点を評価することと、結果が確実であるかどうか評価することを定めておくべきである。以下に述べる重要な因子を考慮する必要がある。

- ・ **増悪日の定義。** 生存期間解析においては、正確な死亡日が判明している。PFS 解析では、正確な増悪日が不明である。以下の 2 つの手法を用いて PFS 解析に使用する記録された増悪日 (P 日) の定義を行うことができる。
 1. 増悪を明言できる最初の日を P 日とする。
 - 増悪日が新しい病変に基づく場合、P 日は新しい病変が検出された最初の観察日。
 - 標的病変測定値の合計に基づく複数の評価が様々な時点で行われている場合、P 日は標的病変測定値の合計が前もって定めた増大を示した最終観察日または標的病変の最終の放射線学的評価日。
 2. すべての放射線学的評価（増悪を集合的に立証する）が行われた直後のプロトコールに定められている来院日としての P 日。
- ・ **打ち切り日の定義。** 打ち切り日は、データのカットオフまたは脱落前に証明される増悪がない患者において定義される。これらの患者における打ち切り日は、増悪状態が適正に評価された最後の日と定義されることが多い。受容できるひとつのアプローチでは、最後の評価実施日としている。しかし、増悪の確認においては複数の放射線学的検査の評価を行う。受容できる第二のアプローチでは、これらの放射線学的評価に対応する来院日としている。
- ・ **適正な PFS 評価の定義。** 増悪エビデンスがない患者において、PFS の打ち切りは、最後の適正な腫瘍評価が行われた日に依存することが多い。適正な腫瘍評価が何によって構成されるかという慎重な定義には、非標的病変の評価と新しい病変検出の両目的のための標的病変評価の適正度と放射線学的検査の適正度の両者を含む。
- ・ **部分的に欠測した腫瘍データの解析。** 解析プランでは、適正な腫瘍評価のための来院によるデータが部分的に欠測している場合の増悪状態を算出するための手法を明記しておくべきである。
- ・ **完全に欠測した腫瘍データ。** 評価のための来院時にデータがまったく収集されていない場合に、その後死亡または評価のための来院で増悪がみられることがある。その他の場合には、後続する評価で増悪がないことがみとめられる。増悪をみとめない場合には、治療継続と増悪モニタリングの継続が適切と思われる。しかし、このアプローチは、後続事象によって欠測データの扱いが異なるため、情報打ち切りに相当する。考えられるもうひとつのアプローチは、後続の PFS 評価によるデータを含めることである。このアプローチは、評価が頻繁に行われ、フォローアップ

本文書は法的拘束力を持たない勧告を含む

プ来院が1回だけ欠測している場合に適している。最後の適正な腫瘍評価時における打ち切りは、来院が2回以上抜けている場合にはより適していると思われる。欠測データによる内在する影響を評価するため、SAPには主要および二次PFS解析について詳しく定めておくべきである。脱落の理由は、打ち切りと増悪状態を検討するための手順に組み込むべきである。例えば、主要解析の場合、文書に記録されていない臨床的増悪、がん治療の変更、または一般状態の低下のために試験を離脱する患者は、最後の適正な腫瘍評価時に打ち切る。二次感受性分析にはこれらの脱落例を増悪事象として含めることになろう。増悪のため欠測になった来院は問題となる可能性があるが、来院の欠測回数とは無関係に、患者の疾患増悪についてフォローアップを継続するためあらゆる努力を払うべきである。

- ・ **測定できない疾患の増悪。** 適切であれば、それぞれの評価モダリティ（例、CT スキャン、骨スキャン）について、増悪基準を説明しておくべきである。測定できない疾患に基づく増悪を立証するスキャンは、盲検判定委員会が証明し、FDAによる確認に供されるべきである。
- ・ **疑わしい病変。** 解析時点において増悪状態を明らかにするため、不確定病変の評価と追跡のためのアルゴリズムを提供しておくべきである。

付録 3 :

PFS 解析のための表の例

III. B で述べたとおり、感受性分析は PFS 解析が確実なものであるかどうか明らかにする際に有用である。しかし、これらの感受性分析は予備的で主要解析結果をサポートするものであり、感受性分析のみに基づいて有効性を主張することはできない。様々な感受性分析を表で説明し、増悪事象の日付および増悪データ打ち切りの日付をどのように定めるかについて指示する。以下の 3 つの表では 3 種類の感受性分析例について説明する。

a. 表 A は、十分立証される明確な増悪事象のみを含む感受性分析である。他のデータは打ち切られている。表 A における増悪日は以下のとおりである：

- ・ IRC により妥当性が証明されている放射線学的評価のみに基づいている。臨床的増悪は増悪エンドポイントとみなされない。
- ・ 腫瘍増悪が記録された最初の日が増悪日と指定される。
- ・ 患者を綿密に追跡していた場合の死亡日。しかし、来院が 2 回以上抜けた後の死亡は最終来院日で打ち切りとする。

表 A. PFS 1 (立証された増悪のみ)

状況	増悪または打ち切り日	アウトカム
ベースラインの腫瘍評価なし	無作為化	打ち切り
予定された来院の間に増悪が立証された	下記の内早い方： <ul style="list-style-type: none"> ・ 新しい病変が明らかになった放射線学的評価日 (増悪が新しい病変に基づいている場合)； または ・ 測定された病変の最後の放射線学的評価日 (増悪が病変測定値の合計の増加に基づいている場合) 	増悪
増悪なし	測定された病変の最後の放射線学的評価日	打ち切り
立証されていない増悪のため治療中止	測定された病変の最後の放射線学的評価日	打ち切り
毒性またはその他の理由による治療中止	測定された病変の最後の放射線学的評価日	打ち切り
新しい抗がん治療開始	測定された病変の最後の放射線学的評価日	打ち切り
最初の PD 評価前に死亡	死亡日	増悪
適正な評価のための来院の間に死亡	死亡日	増悪
1 回以上来院が抜けた後に死亡または増悪	測定された病変の最後の放射線学的評価日	打ち切り

表 B の感受性分析は、打ち切り日と、予定来院日のみにおける事象の日付を明らかにすることによって、腫瘍評価のためのフォローアップ計画において生じうるバイアスを修正する。しかし、このアプ

本文書は法的拘束力を持たない勧告を含む

ローチは、増悪が最終来院日と近接している場合、特にオープン・ラベル試験では、バイアスをもたらさう。このアプローチは盲検無作為化試験に適している。

表 B. PFS 2 (定型的な増悪および評価日)

状況	増悪または打ち切り日	アウトカム
ベースラインの腫瘍評価なし	無作為化	打ち切り
予定された来院の間に増悪が立証された	次回の来院予定日	増悪
増悪なし	適正な評価を行った最終来院日	打ち切り
立証されていない増悪のため治療中止	適正な評価を行った最終来院日	打ち切り
毒性またはその他の理由による治療中止	適正な評価を行った最終来院日	打ち切り
新しい抗がん治療開始	適正な評価を行った最終来院日	打ち切り
最初の PD 評価前に死亡	死亡日	増悪
適正な評価のための来院の間に死亡	死亡日	増悪
2 回以上来院が抜けた後に死亡または増悪	適正な評価を行った最終来院日	打ち切り

- b. 表 C における感受性分析は、治験責任医師の評価に従って PFS を評価する。しかし、このアプローチは、増悪が最終来院日と近接している場合、特にオープン・ラベル試験では、バイアスをもたらさう。このアプローチは盲検無作為化試験に適している。

表 C. PFS 3 (治験責任医師の申し立てを含む)

状況	増悪または打ち切り日	アウトカム
ベースラインの評価なし	無作為化	打ち切り
予定された来院の間に増悪が立証された	次回の来院予定日	増悪
増悪なし	適正な評価を行った最終来院日	打ち切り
治験責任医師が臨床的増悪を主張	来院予定日 (または来院の間であれば次回の来院予定日)	増悪
毒性またはその他の理由による治療中止	適正な評価を行った最終来院日	打ち切り
増悪の主張を伴わない新しい抗がん治療開始	適正な評価を行った最終来院日	打ち切り
最初の PD 評価前に死亡	死亡日	増悪
適正な評価のための来院の間に死亡、または患者が評価のための来院を 1 回行わなかった後の死亡	死亡日	増悪
追跡不能期間延長後に死亡 (評価を 2 回以上受けなかった)	適正な評価を行った最終来院日	打ち切り

付録 4 :
第三者による腫瘍エンドポイントのレビュー

薬剤承認を裏付ける臨床試験結果は、申請者と FDA によって証明されるべきである。単群試験において明らかになった ORR は、限られた数の画像レビューによって評価できる。薬剤承認が PFS 測定に基づいている場合、慎重なプランニングによりバイアスを最小化し、申請者と FDA が結果を証明することができる。IRC は、造影所見の放射線学的解釈と第三者による評価についての判定におけるバイアスを最小化できる。IRC の機能と手順を概説し文書に明記した定款（第三者によるレビュー定款）案は、試験開始前に FDA との合意に達しているべきである。その案には、委員会の独立性の保証と、データの収集、保管、結果の輸送のための手順を記述しておくべきである。定款には、解釈の相違の解決法、データの最終解釈に臨床データを盛り込むこと、および監査手続きも含めておくべきである。IRC の利用については、業界向けガイダンス *Developing Medical Imaging Drug and Biological Products, Part 3: Design, Analysis, and Interpretation of Clinical Studies* において詳しく論じる。