

類似の状況、つまりエビデンスが主に後向きである場合、説得力のあるエビデンスの要件は想定可能である：i) 関連性が強い、ii) 相互作用の生物学的合理性が高い、iii) バイアスを避けるためにデータセット内の被験者の大部分のマーカー状態が判明している、iv) 測定されたアウトカムに対してマーカーの診断性能が許容水準である。時間的關係のような追加要素があるかもしれない。例えば直接トロンビン阻害剤であるキシメラガトランの場合、肝障害は薬剤曝露期間後に進行し、試験継続期間に限定された所定のモニタリングでは、肝障害リスクを検出あるいは軽減できなかつた。よって被験者の適切なフォローアップにより、時間的關係を評価し明確にすることが重要である。この肝障害の遅発は、その後の前向き試験を不可能にしたが、事後（薬剤曝露期間後）の事例を再検討する機会を提供した。

3.4.3. バイアスの低減

バイアスや交絡因子は、GBM の選択と検証に大きな役割を担う可能性がある。これらは、特定の状況にのみ関連する可能性があり、一部が特定の試験デザインだけで注目される。様々な種類のバイアスの中で、選択バイアス及び測定バイアスは、GBM の開発では交絡因子に加えて重要である。バイアスは前向き研究でより最小化しやすく、適切な盲検化やランダム化を含めた試験の適切な設計と実行によって低減されやすい。選択バイアスは後向き解析に著しく影響を与える可能性があるが、特に、すべての関連する試験が利用可能ではなく（出版バイアス）、包括的に記載されない代わりに、試験結果のより都合の良い面に焦点を当てているかもしれないからである。より大きなサンプルサイズは精度を向上させる可能性があるが、バイアスを除去せず、後向き試験に限定されるものではない。後向き試験のための追加の考慮事項には、除外、脱落、及びあるいは報告又は出版バイアスのどれによっても生じる不完全なアウトカムデータによるバイアスが含まれる。測定バイアスは後向き解析での GBM に関連した重要な考慮事項であり、異なる機器や方法が測定に使用されると、特に試験のメタアナリシスでは、共通要素の GBM で起きやすい。明確に定義された測定感度と特異度を有する GBM の集約型測定実験技術や検査は、後向きと前向きの両方で、この低減を助ける可能性がある。更にメタアナリシスと選択用の定義済み基準のある統合されたデータセットを含む試験の慎重な選択も、ある種のバイアスが入り込むのを避け

るのに役立つ。

3.4.4. 多重性

試験の前向きあるいは後向きに関係なく、多重性の問題（多重比較による偽陽性の誤判別率の増加）は、GBM 開発において対処する必要がある。この観点での多重性は2つの異なる面を含んでいる。1つはアウトカムと十分に強い関連性を持っていることを特定しようとする多重 GBM 又はパネルの使用である。潜在的な多重 GBM を開発プログラムで検討する場合、調べる GBM の数は、全ゲノムでの潜在的関連性を検討できるシグナル生成アプローチであるか、あるいは探索の根拠がより絞り込まれた場合のより少ない潜在性 GBM になるのかに依存している^{17, 18)}。これらの問題は、単一 GBM では十分な予測ができず、多重検査 [multiplex testing] に利点があるかもしれない一般的な多因子遺伝病において、より重要と見ている。ここでの多重性制御の主な目的は、企業や研究者が信頼性の高い手がかりのみを追跡することと、企業と規制当局の双方による特定された関連性の強さのエビデンス評価である。

2つ目は臨床試験中の多重検定周辺の問題である。前向き研究で検討する GBM の場合、スポンサーは解析計画での多重検定周辺の問題の検討を望み、適切に遂行されていれば、多重検定による規制上のリスクを管理する必要がある。後向き評価では、形式上制御することができず、そのため推論は特に慎重にしなければならない。しかし、多数の可能な補正が、Bonferroni, Benjamini-Hochberg, Sime の論文を含む文献で提案されている。方法論的観点から、遺伝的相互作用に対する多重検定の相関的性質に対応しつつ、有意性の虚偽の主張から保護する統計手法は合理的である。Bonferroni の補正は保守的との批判があるが、規制の観点から、より厳格な補正法の下で統計的有意性を保持している関連性は、特にエビデンスが後向き試験から得られたものであるとき、より説得力があるかもしれない。多重性の問題に関する CHMP のガイドライン (CHMP/EWP/908/99) も参考になる。

17) Yang Q, Khoury MJ, Botto L *et al.*: *Am J Human Genet.* 72 : 636-649, 2003.

18) Janssens AC, Pardo MC *et al.*: *Am J Hum Genet.* 74:585-588, 2004

臨床開発と患者選択に関連したゲノム薬理学バイオマーカーに伴う方法論的問題への考察書 草案^{*2)} <2>

EMA ヒト用医薬品委員会

Reflection Paper on Methodological Issues Associated with Pharmacogenomic Biomarkers in Relation to Clinical Development and Patient Selection ; Draft <2>
(9 June 2011, EMA/446337/2011)

European Medicines Agency Committee for Medicinal Products for Human Use (CHMP)

翻訳：嶋澤 るみ子^{*1, #}, 池田 正行^{*1}

4. GBM の開発

4.1. 探索的開発段階

4.1.1. 非ランダム化[コホート, 症例対照又は単一群]試験

GBM は、非ランダム化コホート又は単一群試験の探索的パラメータとして、医薬品開発プログラムの内外で頻繁に用いられている。これら GBM は、疾患の重症度、転帰などに対する予後性のものかあるいは、単一又は組み合わせ療法に対する特定応答の予測性のものかもしれない。GBM の同定と開発のためのこのような研究は、特に初期の段階は試験デザインが大きく異なる可能性が高い。これら探索的研究は、選別が不完全で、サンプルサイズが限定された便宜的コホートの傾向があり、多くの場合、GBM の予測値を確立し、その感度と特異性を定量化する厳密さを欠いている。多くの研究が、明確に立証され事前定義されたバイオマーカーに関連するエンドポイントや解析計画を欠いている。一部の研究では、

適格基準が組み入れ時のバイオマーカーの状態と無関係な可能性がある。これは非選択 [unselected] 試験デザインに相当し、その長所があるかもしれないが、GBM に基づく治療割付けがないのが限界となり、マーカーの真の検証とはならない。

ゲノムワイド関連解析 (GWAS) 一関連性の横断的試験一は、表現型に大きな多様性が存在するが、興味深い共通の特徴がある場合、しばしばゲノムマーカー同定に有用な手段として機能する。これは特定の開発デザインとしてより、検索の戦略として機能する。後向き関連解析 (GWAS) が、GBM、疾患、薬物応答間の関係について初期エビデンスを提供するとき、しばしば上記・非選択試験デザインと同様に真の検証とされない制限を受ける。あらゆる後向き解析において、関連性が選択バイアスと、GBM で特定された集団の表現に影響する他の一般的な潜在バイアス回避のために、大部分の被験者からの生物学的サンプル (血液、組織他) が利用できることで主に成立する場合母集団 (サンプル) サイズを考慮す

^{*1} 長崎大学大学院医歯薬学総合研究科生命医科学講座創薬科学 長崎市坂本 1-12-4 (〒 852-8523)

Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

^{*2} 本翻訳は厚生労働科学研究費 (医薬品・医療機器等レギュラトリーサイエンス総合研究) 「コンパニオン体外診断用医薬品の臨床性能試験の在り方に関する再帰的研究」として、他国のコンパニオン診断薬の規制要件状況調査を目的としたものである。(※コメントの提出方法等の部分は省略している。また出典の更新と訂正を行っている。) <1>: 医薬品医療機器レギュラトリーサイエンス, 43(11), 1010～1016.

[#] 責任著者 Corresponding author

ることが重要である (3.4.3 項及び 4.2.2 項参照)。

症例対照研究で症例数が限られている場合、有用な情報が提供されることがあるが、症例と対照群が由来する集団全体が十分に多数であっても、決定的エビデンスは得られないかもしれない。症例対照研究で考慮する主な点は、症例と対照群に適用する定義や知見を一般集団へ外挿できること、治療的介入を含む2つの集団の扱いに存在するすべての違いである。GBMを用いる疾患定義や特定の治療に伴うリスクには、選択バイアスの可能性がある (すべての後向き操作に当てはまる)。症例対照研究は、治療的介入又はその割り当ての有用性と、介入の自然経過への真の、バイアスのない影響を判断する能力を限定するかもしれない後向き評価である (例えば、乳がんにおけるタモキシフェンに関する議論、遺伝子型に基づくワルファリン投与)。症例対照研究とは対照的に、コホート研究は前向き又は後向きが可能であり、疾患の発生率と自然経過に関する情報を提供するが、同時対照群がないため薬物反応が乏しい可能性がある。このコホート研究の重要な一面は、患者の選択がマーカーではなく、他の臨床パラメータに基づいているかもしれないことである。薬物応答の予測性 GBM 開発では限られた価値しかないかもしれないが、定義されたアウトカムを伴う興味あるマーカーへの手がかりを提供するかもしれない。しかし、これは外的影響や交絡変数により制限される。ランダム化アウトカム試験を背景とした SLC01B1 多型、高用量スタチン使用とミオパシー間の関連を評価する GWAS (SEARCH 試験¹⁹⁾) は、興味深く、交絡因子のいくつかを強調している。

時々、GBMに関する予備的情報は、同じ種類の他の薬剤又は特性が共通な薬剤のそれまでの所見に起因する可能性がある (例えば、CYP2D6 低代謝群 (PM) における有害事象率の増加が、CYP2D6 基質の薬剤分類全体に及ぶ可能性がある)。したがって、新しい薬剤にとって、薬剤の登録又は承認に先立ち、ヒトでの特定の GBM の相対的な重要性の確認を、早期に得ることが適切である (例えば、新しい CYP2D6 基質薬剤の効果における CYP2D6 遺伝子多形の役割)。GBM に関するデータが、登録又は承認後、あるいは特許期間終了後に入手された場合、その後臨床試験が標的集団で計画、実行される必要があるかもしれない。このような開発プログラムは、コホート研究と前向き RCT の両方が関与する可能性が高い。この状況でのコホート研究は、マーカーの背景情報を提供する可能性があり、前向き RCT は、交絡変数

目次

1. 序論
 2. 範囲と目的
 3. ゲノムバイオマーカー (GBM) の特徴
 - 3.1. GBM の分類
 - 3.1.1. 予測性 GBM
 - 3.1.2. 予後性 GBM
 - 3.2. GBM の選択
 - 3.3. GBM の目的
 - 3.3.1. 患者選択
 - 3.3.2. 治療アルゴリズムの割付
 - 3.4. GBM 特有の考慮事項
 - 3.4.1. 特定種類の GBM に対する技術的考慮すべき事項
 - 3.4.2. シグナル出現のタイミングと臨床開発に対する影響
 - 3.4.3. バイアスの低減
 - 3.4.4. 多重性
- 上記は本誌11月号掲載 (Vol.43, No.11)—
4. GBM の開発
 - 4.1. 探索的開発
 - 4.1.1. 非ランダム化 [コホート、症例対照又は単一群]試験
 - 4.1.2. ランダム化対照試験 (RCT-前向き又は後向き評価)
 - 4.2. 検証的開発段階
 - 4.2.1. 前向き検証のための試験デザイン
 - 4.2.2. 異なるデザインの比較 (長所と短所) 後向き検証は可能か? (確認)
 5. マーカーの診断性能
 - 5.1. 感度, 特異度, NPV (陰性適中率), PPV (陽性適中率)
 6. GBM 評価のための機器/診断キット
 7. GBM 評価に関する潜在的な外部からの影響
 8. 他の側面
 9. 用語集

の影響を減らして真の効果を評価する。CYP2D6 遺伝子多型の役割とタモキシフェンの使用について進行中の議論は、データが市販後段階で利用可能になった場合の困難さを強調する事例である²⁰⁾。Schroth らは、1325 患者

19) The SEARCH Collaborative Group—A Genomewide Study *NEJM*. 359 : 789-799; Aug 21, 2008.

20) Schroth W, et al. : *JAMA*, 2009, 302 (13) : 1429-36.

21) Wegman P, et al. : *Br Cancer Res*, 2005, 7 (3) : R 284-90.

の後向きコホート研究で CYP2D6 遺伝子多型の影響を検討し、Wegman らは²¹⁾、680 患者の集団の約 220 被験者でこれを評価した。これらの研究は、組入患者の状況、検討された治療法、遺伝子型解析のための腫瘍組織入手の可能性が異なっている。他の研究²²⁾では、追加の GBM を評価し、マーカー間の相互作用と後向きにマーカーの重要性を評価する複雑さを強調している。

4.1.2. ランダム化比較試験 (RCT-前向き又は後向き評価)

ランダム化臨床試験による GBM の探索的研究 (仮説生成) は、公表された文献や開発プログラム内の初期試験に基づく予測性 GBM の値に関する予備的情報から得られることが多い。これらは、新たな前向き RCT、あるいは完了した臨床試験からのデータの後向き解析の可能性がある。GBM の同定と検証のための前向き RCT の活用は、理想的だがいくらか制約がある；高価で、時間又は労力集中的であり、多くの場合 RCT に先だち関連性又は生物学的な合理性のどちらかを実証する十分な予備的エビデンスを必要とする。この場合に適用されるデザインは、検証のためのピボタル試験と同様であり、本文書の 4.2 項で議論されている。あるいは完了した 2 つの異なる薬剤又は治療戦略を比較している RCT の後向き解析は、仮説生成の機能を果たすかもしれない。

そのような後向き探索や検証のために、以下の一定の要素が必須である：データは適切に実施された RCT から入手できる必要があり、選択バイアスを避けるために、臨床試験に十分な多数の被験者からの GBM データが入手できる必要があり、解析計画は事前に規定する必要がある。パニツムマブでの経験は代表例である。ピボタル第 III 相試験では、EGFR の状態が選択基準であり、したがって、GBM データはすべてのランダム化被験者で利用可能であり、選択バイアスの可能性は減少し、探索的検討としてではあるが、KRAS 変異による解析が事前規定されていた。2 つ以上の独立し、適切に実施された RCT から得られたデータの解析は、最強のエビデンスをもたらす。ほとんどの場合、1 つのピボタル RCT から補強エビデンスが得られることが予測される。

4.2. 検証の開発段階

GBM の役割を確立するための検証的段階は、単一 GBM や GBM 識別特性 (GBM のパネル) が、開発の初期段階で臨床的妥当性の取得へ進むのに十分な厳密さで有望であることを前提とする。探索研究での高い正負の予測値を伴う GBM は関心の 1 つだが、選択に適用する

厳密性の水準はその都度決定する必要があり、ここで特定することはできない。

4.2.1. 前向き検証のための試験デザイン

検証の開発で使用される試験デザインは、マーカー間の異なる要因によって影響される可能性が高い。例えば、関与する経路又はマーカー；マーカー、疾患及び計画された介入間のメカニズム又は生物学的関係；集団でのマーカー出現率と遺伝パターン；仮定された効果の大きさ；民族、性別の影響；利用する層別化を含む計画された解析などである。試験開始時に、GBM の使用可能な分析的妥当性も重要な要素になる可能性が高い。

RCT は前述したように、バイオマーカー (特に予測性マーカー、本文書の主な焦点) の前向き検証としてのピボタル/検証的試験の推奨デザインである。数種の RCT—非選択、強化 [enriched] 又は標的 [targeted]、ハイブリッド [hybrid] とアダプティブ (適応的) デザイン—が可能であり、後者 3 つは組み入れ集団と最終的な解析に関してより具体的である。以下に、デザインのいくつかの側面を説明している。マーカー出現率が稀な場合、後述のシナリオはどれも最適ではないかもしれないので、追加の助言を得るのが最善である。

非選択デザイン RCT

一般に、非選択デザインを用いる試験は、試験の組み入れ適格性がバイオマーカーの状態に基づかない場合に最も有用である可能性が高い。非選択 RCT は、a) 逐次検定戦略 [sequential testing strategy] デザイン、b) マーカーベース [marker-based] デザイン、又は c) ハイブリッドデザインに大きく分類でき、プロトコルに記載された手段によって互いに区別される。主要解析は採択された戦略に依存する。例えば、逐次戦略デザインでは、集団全体での治療への応答が主要解析で、副次解析がマーカー依存性応答かもしれないが、この解析計画は変更、つまり主要解析としてのマーカー依存性応答と、副次解析としての全体集団の応答が可能である。このようなデザインのサンプルサイズ要求は、他のデザインよりも大きくなる可能性があり、事前に定めた GBM に基づく解析で、ベネフィットを明確に実証することが期待される。これは、試験全体では明確なベネフィットが示されていないが、GBM に基づく解析で示している場合に、意思決定に困難をもたらす。これは、GBM に基づく解析を事前に定めることで克服可能かもしれない。このような状況では、1 つのピボタル試験に基づく申請要件を考慮することが重要である。しばしば副次的な GBM ベース解析の結果は、十分な検出力のある他のデザインを使える 2 度目の試験で、更に確認を必要とする可能性

22) Kiyatoni K, et al. : *J Clin Oncol*, 2010, 28 (8) : 1287-93.

が高い。転移性結腸直腸癌のサードラインにおけるパニツムマブ使用を評価する 20020408 試験では、主要選択基準であるすべての EGFR 陽性腫瘍を有する被験者を募集するのに、このデザインを使用していた。野生型 KRAS 又は変異型 KRAS に対する応答は、事前に計画された副次又は探索的解析であった。試験への被験者選択は KRAS 変異の状況に関連しておらず、この野生型対変異型 KRAS の解析のためであり、試験は非選択デザイン試験として機能した。

強化デザイン RCT (標的デザイン)

強化又は標的デザイン (Fig. 1) では、マーカーの状態が重要な適格基準、つまり被験者はマーカーの有無に基づいて組み入れられる。強化又は標的デザイン RCT を用いるには、GBM と疾患を結ぶ強力な生物学的合理性と、GBM と薬物応答間の関連性に説得力のある予備的エビデンスが必要である。これは GBM が試験集団を定義するので、GBM 定義集団以外の被験者の除外理由は、明確に定義する必要がある。GBM が直接治療(薬物)標的を構成するか、影響を与える場合、標的強化デザインが最も適切である。強化デザイン試験で最も評判の良く成功した例は、ドキシソルピシンとシクロホスファミド併用療法後の HER2 陽性術後患者におけるパクリタセル併用のトラスツズマブへの応答を評価する試験であり、トラスツズマブは、DFS (無病生存) ハザード比を約 25%減少させた。

強化デザインは、分析の正確度と再現性が非常に良く確立され、臨床試験の完全性と問題となっている実際のベネフィットを損なうおそれがある被験者の GBM 陽性又は GBM 陰性の誤分類の機会、あるいは可能性がほとんどないことを前提としている (3.4.1 項を参照)^{23,24)}。このデザインは、その多くの型で、マーカーで定義された治療へのランダム割付け群でのアウトカム差の検出力が低く、疾病集団の残りの部分に関する情報は何も得られない。したがって、マーカーに基づき選択された集団での治療のベネフィットリスク比のみを検証する。このデザインは、集団全体での治療効果は中程度だが、GBM 陽性又は GBM 陰性のマーカーで定義された集団において許容できない水準のリスクを伴う場合に、最も価値のある可能性が高い。被験薬とプラセボ間の奏効率の差が、プラセボ群の GBM 陽性と GBM 陰性の被験者間の奏効率の差と同様である場合、別に評価されていた場合でも、強化デザイン試験の予測値は、役に立たない

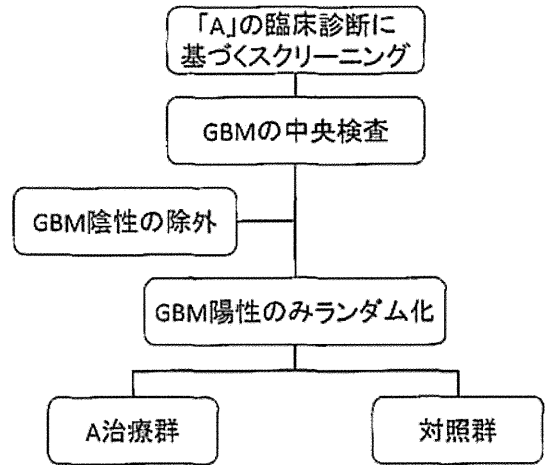


Fig.1 強化/標的デザイン

可能性があることに注意することが重要である。

一般に強化デザインは、非選択集団では、治療のベネフィットが中程度で重篤な毒性が伴う場合や、非選択デザインが倫理的に不可能かもしれないときに最も有用である。

マーカーベースデザイン

マーカーベースデザインは、2値マーカーとの関連で医薬品開発や検証のために採用された例が数多くある。これらの例は、マーカー・治療交互作用 [marker by treatment interaction] デザイン、あるいはマーカーベース戦略デザインである可能性があるマーカー・治療交互作用デザインでは、層別化の手段としてマーカーを用い、患者は各部分集団内での治療に割り付けられる (Fig. 2a)。これの主な利点は、サンプルサイズが各部分集団内で事前に定義され、2つの RCT と同等になることである。

マーカーベース戦略デザインでは、患者がマーカーの状態に基づきあるいは無関係にいずれかにランダムに割

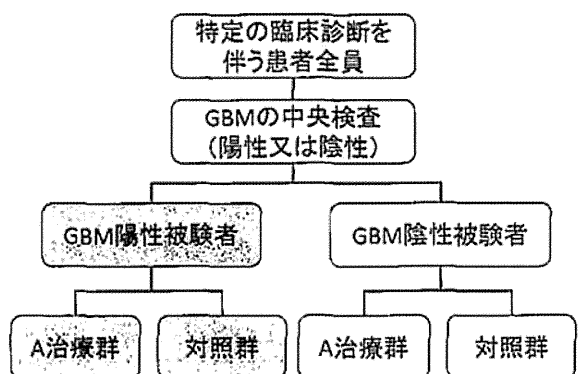


Fig.2a マーカー・治療交互作用デザイン

23) Perez EA, et al. : J Clin Oncol, 2006, 24: 3032-3038.

24) Paik S et al. : J Natl Cancer Inst, 2002, 94:852-854.

り付けられる。Fig. 2aには示されていない後者の場合は、全体の検出可能なアウトカムの差が減少し、サンプルサイズがより大きくなる。

ハイブリッドデザインのRCT

ハイブリッドデザインでは、GBM定義された被験者の部分集団のみランダムに、マーカーの状態に基づいた試験対象の治療に割り付けられるが、他のGBM定義された被験者は標準治療に割り付けられる (Fig. 2b)。この試験は強化デザインに似た検出力があるが、この戦略は付加価値を加えているかもしれない。このデザインは治療が複数の薬剤、あるいはある治療法での有効性の有力なエビデンスがある戦略の場合に最も有用である。標準治療は、それゆえ事前に定義されている治療法が含まれ、実験群と全体的な試験は、GBM定義された被験者に、その後の又は追加治療の選択肢に関する追加情報を提供するだろう。

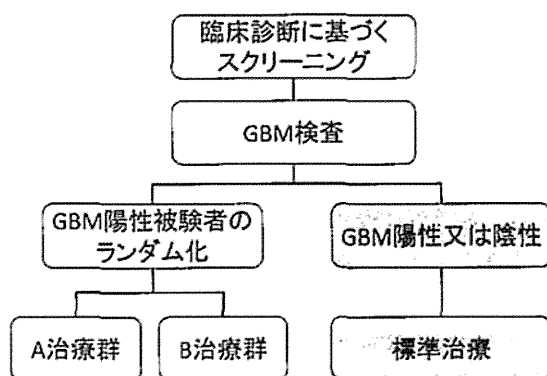


Fig. 2b ハイブリッドデザインの一例

ハイブリッドデザインの利点として認識されているのは、標準治療が終わってからの増分効果と事後の比較の可能性であることだが、スクリーニングで得られたGBM検査用サンプルが、他の予後性マーカーの将来の検査のために保存されている必要がある。ハイブリッドデザインの変更版は、アバカビル過敏症を評価する Predict-1 試験で使用されていた。HIV の臨床診断を受けた被験者は、遺伝子検査群と標準治療群にランダムに割り付けられた。前者の群では、HLA-B* 5701 陽性患者を除き、検査陰性のみアバカビルが投与され、標準治療群は、アバカビル投与前には HLA-B* 5701 検査は行われなかった。両群の過敏反応がモニターされた。これは安全性相互作用の古典的事例だが、HLA-B* 5701 マーカーの有用性の評価試験と解釈される可能性もある。

アダプティブ (適応的) デザイン

近年アダプティブデザインへの関心が高まっている。これらは、適応的閾値 [adaptive threshold] (統計解析) デザイン、適応的増加 [adaptive accrual] デザイン、あるいはアウトカムに基づく適応的ランダム化 [adaptive randomisation] の可能性がある。これらの組み合わせも可能である。適応的閾値デザインは2つの解析法が可能である；第1は α 消費に依存する事後比較のための異なる有意水準と、事前に閾値を決定した有意水準の全体比較であり、第2はGBM陽性被験者と検査についてのみ有効性を仮定した解析である。このようなデザインは、治療効果の検定と選択したマーカーのカットオフ点の前向き検証が必要なきも有用である。適応的ランダム化の計画は、無益性の中間検定に基づき、特定の治療群への修正増加を可能にする。現時点で入手可能な成功したアダプティブデザインの例は規制上の意味ではほとんどなく、実際にこのようなデザインは臨床試験でほとんど試されていないが、可能性はある。

4.2.2. 異なるデザインの比較 (長所と短所)

GBM の検証に使用する臨床試験の異なるデザインを比較する豊富な文献がある。入手可能なエビデンスと科学的背景に基づいて、高い生物学的合理性を伴う真の予測マーカーがある場合には、強化デザインが最も効率的である可能性が高いが、2つの要件を満たす必要がある。第1は、マーカーの状態を決定するためのカットオフ点があること、第2は、臨床試験の価値がなくなるのを避けるために、誤分類が起きない保証の必要があることである。マーカーのカットオフ点確立していないが、マーカー出現率が高い場合には、非選択デザイン又はその改変が最適な選択肢になるが、より大きなサンプルサイズの必要性に悩まされるかもしれない。それに引き換え、分析の正確度、再現性及びマーカー出現率に依存しているとはいえ、標的デザインは非選択デザインと比較して、より少ないランダム割付患者、そして実際より少ないスクリーニング患者で済む。臨床試験からGBM陰性患者を除外することが規制上許容されるか否かは、これらの患者での効果の欠如を示す証拠の強さ (妥当性、科学的根拠、臨床データ) に依存する。

強化デザインの限界：強化デザインはGBM自体ではなく、特定の集団で問題となっている治療の利点のみを検証する。GBMに基づく強化デザイン試験で示された実薬・プラセボの差が、プラセボ治療を受けたGBM陽性とGBM陰性被験者の差と同じであるかは、結果とは無関係かもしれない。GBM陽性又はGBM陰性被験者の分類に使用される分析正確度は非選択デザインと標的デザインでは異なる影響を与える。誤分類がある場合は、

非選択試験ではマーカーに関してだけ影響を与える可能性があるが、強化試験では、これはマーカーに関する推論に加えて、全体の完全性と試験の結果を危うくするかもしれない。

マーカーベースデザインは、特定の状況で利点がある。2値分類を具体化できる2値マーカー又は多重マーカーの識別特性において、より小さなサンプルサイズと高いイベント発生率あるいは群間でイベント発生率のより大きな差は、非選択デザインに比べて、マーカー・治療相互作用デザインで可能性が高い。マーカーベース戦略デザインは、潜在的な不利マーカーベース群と非マーカーベース群の両方で同じレジメンで治療を受ける患者の重複がある。注意点は、癌治療分野での経験がほとんどで、他分野での適用性は検証されていないことである。

後向き検証は可能か？（検証）

新たな前向きデザインの試験が様々な理由で実行できない場合、これまでに適切に実施された治療法を比較するRCTのデータを利用して、マーカーの予測能力を検討する可能性を特定の状況で考慮できる（後向き検証）。どの後向き検証にとっても、1つ以上の適切に実施された前向きRCTのデータ、そして多数の被験者からGBM状態を入手できるなどの決定的要素が、選択バイアスを避けるために重要である。更に、検証する仮説と解析計画は、後向き評価の開始前に文書化する必要がある。前述のように後向き検証では、1つ以上の独立したデータソース又はRCTの使用により必要なエビデンスの提供ができる。後向き検証に含まれる試験のデザインは、前向きを検証試験に類似する可能性が高いが、GBMに関しては非選択デザインが圧倒的多数の可能性が高い。2つの方法の違いの1つは、既に完了したRCTの後向き解析では、試験組み入れの適格性がマーカーの状態に基づかない、つまり非選択デザインの可能性がある。これはマーカーの検証の助けになるかもしれない。転移性直腸結腸癌の野生型KRASとパニツムマブ投与後の無増悪生存期間（PFS）の改善との関係を確認する研究は、後向き検証の一例である²⁵⁾。この事例では、野生型と変異型KRASの保有者間でのパニツムマブの効果の差が、GBMの事後解析により示され、セツキシマブの臨床試験から得られた関連性の生物学的合理性と

もに、欧州での条件付き承認の基となっている。承認は、前向きに作られる更なるデータを条件として要求していた。GBMの状態の情報が、被験者の大部分から得られる必要があるとの考慮事項は、20020408試験で満たされていた。この1つのRCTから得られた実証データで、解析が事前に定義され、選択バイアスを回避して被験者の大部分のGBM状態が、決定可能であった。

それに比べて、又は対照的に、ゲフィチニブでのEGFR FISHと/又はEGFR変異状態間の相互作用は、いくつかの試験で評価されていたが、非小細胞肺癌患者での探索が目的であった。これらの試験には、多様な患者集団（アジア人でのISEL試験）、希望者全員対象のINTEREST試験（コーカサス人種が圧倒的多数）、及び混合集団でのIPASS試験が含まれている。これらの試験は前向きであったが、INTEREST試験のみEGFR FISH陽性に基づく違いを、共同主目的 [co-primary objective] として事後（後向き）解析を効果的にしていた。これらの試験で示されている応答率の差は、民族性、他の臨床的特徴あるいは前治療の違いによって影響されているかもしれない。マーカーの状態が既知/同定された被験者数の違いが、それぞれのバイオマーカー（EGFR FISHの状態、EGFR変異状態、EGFR蛋白の発現）において、役割を持っていた可能性もある。全く異なる結果にもかかわらず、様々な比較間での方向性の一致と、EGFR変異状態とゲフィチニブへの応答の間で再現された相互作用のために、統合解析は、EGFR変異陽性腫瘍の場合のみゲフィチニブ治療のベネフィットを示した。特に、これら複数の試験と統合解析の結果に基づいて、CHMPと専門家助言集団の双方が、非小細胞肺癌でのゲフィチニブの広範な効能に同意せず、ゲフィチニブへの応答がEGFR変異状態に影響されている、制限された効能を承認すると判断した。ゲフィチニブ開発プログラムへの批判の1つは、様々な試験での全被験者でのバイオマーカーに関する情報が欠如していることで、これはより良く設計、計画することができるものだった。この例はGBMの後向き評価の2つの重要な側面—第1に様々な試験や集団におけるGBM・薬物応答の相互作用の再現と、第2に解析において全被験者からのGBM状態の情報を最大にする必要性—を強調している。

以上から、次の点が満たされていれば、規制上/科学的な観点で、後向き検証や後向きデータの受け入れが可能かもしれない。実施されたRCTからのデータ；RCTにおける被験者の大多数のマーカー状態の情報が利用可能；仮説と解析計画が事前定義されていること；多重検定で調整された統計的に説得力のある関連性；そして最終的に1つ以上の独立したサンプルでの結果の再現、で

25) Vectibix EPAR-2007.
http://www.emea.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion/human/000741/WC500047707.pdf.

ある。

5. マーカーの診断性能

GBM あるいはバイオマーカーの診断性能の特性を明らかにし、この性能が疾患の特徴、他の治療的介入の効果、GBM 発現の性別や民族による疫学的な違いにより影響を受けることを説明するのは重要である。例えば、HLA-B*1502 と抗てんかん剤曝露によるスティーブンジョンソン症候群の関係は、民族的多様性を示し、広い非選択集団で使用する場合、性能に影響する可能性がある。

診断性能の基準は、適格性評価とバリデーションの一般的な基準に適合する必要がある。反復性、再現性と精度推定等の特定の側面を評価する必要がある²⁶⁾。診断性能は感度と特異度の評価に基本的に関係している。

GBM 又は他の従来のマーカーの値は、大抵検査される集団に特有であり、実用性は臨床使用を反映した集団で実証される必要がある。他の集団へのどの外挿も、安全性 GBM のリスクの推定基準を含めて、検査・外挿された集団で実証された表現型、GBM の出現率を考慮し、適切に正当性が証明される必要がある。

以下で取り上げた事項は、一般にマーカーと一部の検査の事例の双方に適用可能であるが、診断薬の詳細な議論は本文書の範囲を超える。

5.1. 感度、特異度、NPV (陰性適中率)、PPV (陽性適中率)

探索で使用する試験の種類に関係なく、予測可能性と臨床的妥当性に関するバイオマーカー性能の情報収集は重要である。マーカーの臨床的妥当性とは、検査/マーカーの感度と特異度とゲノムの異常又は変異の浸透率の複雑な相互作用である。開発的な研究は仮説生成に加えて、GBM の更なる評価が実行可能で価値があることを示すことを目指すべきである。この点は第 II 相臨床試験と同等の役割だが、コホート又は単一群試験における探索的データセットの限界により影響を受ける可能性が高い。これらの試験で得られたデータの検証が何らかの形で得られることが期待され、GBM 自体の出現に使用していないデータで実行される必要がある。そのプロセスは GBM の予測性能に係る情報を収集すべきで、これは検証段階で更に評価される正負の予測値を含める必要

がある。

PPV 及び NPV は、マーカー出現率と対象の集団での表現型発現に依存する可能性がある。探索的研究で高い正負の予測値を伴う GBM は特に関心を持たれるだろうが、適用される厳密性の水準が、特に規制の観点では不明であり正確に規定できない。95%の特異度さえ許容できない場合があり、国立がん研究所は、有病率の低い卵巣癌のスクリーニングを例にしている²⁷⁾。マーカーの最終的な評価は、特に他のマーカーや特性が利用可能である場合は大抵多変量解析である。多重マーカーのパネル又は識別特性が利用されている場合、多変量解析の結果は、一定の制限—GBM で採用したカットオフ値、連続あるいは分類か、他の既存のマーカーがすべて適切にコード化されているか、変数がどのようにモデル化されたか—に依存する。尤度比、一致指数あるいは一致指数の変化に基づく代替方法が提案されており、状況によっては適切かもしれない。Yang らは、尤度比を、疾患を有する患者が実際の検査結果を有する確率と定義して用い、疾患のない患者が同じ結果を有する確率と比較した¹⁷⁾。一致指数はランダムに選択された 2 患者のうち、より予後が悪い者が、実際により悪い予後を予測される確率である²⁸⁾。一致指数は、受信者動作特性 (ROC) 曲線下面積と類似している。

0 と 1 又は確定したカットオフ値の上下のような値をもつ 2 値モデルでは、ROC 曲線が、階層分布とは独立した最適なモデルを選択する手段を与える。ROC 曲線は、マーカーの数値が関連する連続データセット内の別のカットポイントの使用も可能にし、任意の点における傾きから尤度比を計算する機能を提供し、曲線の曲線下面積が検査の正確度の尺度になる。予測コンソーシアムで評価された腎損傷マーカーでは、いくつかのマーカー (KIM-1, アルブミン, シスタチン-C など) で、この利用に成功しているが²⁹⁾、規制上の観点からは、GBM のこの申請成功例はほとんどない。

6. GBM 評価のための機器/診断キット

GBM 及び他の疾患や状況で GBM の検出に関連した分析は、市場で入手できる場合があるが、この場合の検査の検証は、既に利用可能かもしれない許容できる分析や方法として評価段階にある疾患と治療に限定される

26) Biomarker qualification—renal biomarkers.
http://www.ema.europa.eu/pdfs/human/biomarkers/28329810_en.pdf.

27) <http://www.cancer.gov/cancertopics/understandingcancer/moleculardiagnosics/page34>.

28) Kattan MW. : *Clin Cancer Res*.10: 822-824. 2004.

29) Nephrotoxicity BMs. Final conclusions of EMA/FDA VXDS experience. EMEA/679719/2008 Rev.1

(例えば HLA 対立遺伝子)。新たに同定された又は特定の GBM は、医薬品開発と併行して、特定の分析/キットの開発が必要かもしれない。ほとんどは薬剤と「コンパニオン診断薬」の共同開発であり、市販検査の分析特異度と感度に関する問題を考慮する必要がある。同様の考察は、GBM 検出に利用できる他の分析又は検査（一般的に「自家製の検査」と呼ばれるものなど）に適用される。本文書の目的は、主に GBM に関連する方法論的側面を議論することであり、市販検査やキットの性能ではない。しかし、以下のような見解がある。

- ・ 医薬品開発の観点から、開発プログラム内にコンパニオン診断薬が含まれる可能性がある。特定の診断キットや方法論をピボタル試験で採用し、その検査が GBM の同定/定量に特有である場合、その特定の検査方法と GBM の値を関連づける必要があるかもしれない。免疫組織化学 (IHC) 又は FISH 検査を利用した HER2 過剰発現の同定と定量はその一例である。この例では、医薬品のラベルが、この追加情報を結果の記述の一部として提供する。他にもこのような例が存在する (Dako 社製検査とセツキシマブ, Monogram 社 trofile 検査とマラビロク)。
- ・ GBM 同定に使用する検査がより包括的性質、すなわち特別に開発した GBM ではない場合 (例えば、CYP2D6 遺伝子多型の同定)、コンパニオン診断用検査の特性に関する議論が、医薬品ラベルに含まれる可能性は低い。特定の診断キットの提供も期待されない。
- ・ 臨床試験で使用した分析/検査が利用できず、新しい又は別の検査に代える可能性がある。この場合、GBM マーカーの状態を測定する臨床試験の分析と市販の検査が同等であることが期待される。同様の他の市販検査間の同等性も重要な検討事項であるが、これらの議論は本文書の現在の範囲を超えている。

7. GBM 評価に関する潜在的な外部からの影響

マーカーと治療効果の関係は、表現型と関連し、問題となっている薬剤への個々の応答を決定している他のゲノム、又は非ゲノムの要因によっても影響される可能性があることを認識すべきである。これは検討している標的疾患あるいは健康状態とは独立しているかもしれない。例えば、パロキセチンとタモキシフェンの同時投与や併用のような CYP450 基質又は阻害剤の使用は、ピボタル試験で制御されていない場合、ER 陽性乳癌の治療結果に影響するとの議論がある。これはタモキシフェ

ンの代謝 (CYP2D6 基質) に影響を与える特定の CYP2D6 遺伝子多型による効果と似ている。

同様に、薬物受容体に関わる多型は、基礎疾患あるいは疾患とは独立に、1つ以上の薬物への応答に影響を与える可能性がある。薬剤への応答は、他の代謝酵素の遺伝子多型によっても影響を受ける可能性がある。例えば、ワルファリンに対する応答での CYP2C9 遺伝子多型の相互作用は、同時に VKORC1 遺伝子多型の遺伝的形質によって影響される。したがって全体的な効果は、2つの影響の複合であり、評価中には相対的な寄与に関して明確にすることを要求すべきである。

初期研究におけるゲノム薬理学 GBM 評価の詳細な議論は、別文書 (EMA/CHMP/37646/2009-下記参照) が利用可能である。

8. 他の側面

本考察書は、以下のガイダンスと併せて読まれるべきである。

- ・ Position paper on terminology in Pharmacogenetics (EMA/CPMP/3070/01)
(薬理遺伝学の専門用語に関する方針説明)
- ・ Reflection paper on pharmacogenomic samples and data handling (EMA/CHMP/201914) (ゲノム薬理学のサンプル及びデータ処理に関する考察書)
- ・ ICH Topic E 15 : Establish definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories (CHMP/ICH/437986/2006)
(ICH トピック E 15 : ゲノムバイオマーカー, ゲノム薬理学, 薬理遺伝学, ゲノムデータとサンプルコード化分類の定義の確立)
- ・ Reflection paper on the use of pharmacogenetic in the pharmacokinetic evaluation of medicinal products (EMA/641698/2008) (医薬品の薬物動学的評価における薬理遺伝学の利用に関する考察書)
- ・ EMA/CHMP/SAWP/72894/2008 Corrl. Qualification of Novel methodologies for Drug Development: Guidance to Applicants. (医薬品開発のための新しい方法論の適格性: 申請者向けガイダンス)
- ・ EMA/CHMP/ICH/380636/2009; Genomic biomarkers related to drug response: context, structure and format of qualification submissions. (薬剤応答に関連するゲノムバイオマーカー: 適格性に関する提出の背景, 構成と形式)
- ・ EMA/CHMP/37646/2009; Guideline on the use of

pharmacogenetic methodologies in the pharmacokinetic evaluation of medicinal products. (医薬品の薬物動態学的評価における薬理遺伝学的手法の利用に関するガイドライン)

9. 用語集

Accuracy：正確度

測定値とその真の値への近さの示す尺度

Analytical Validity：分析的妥当性

ゲノムバイオマーカーにおける分析的妥当性とは、対象の遺伝子型（マーカー）を正確に（確実に）測定する特定の検査能力を説明するもの。これは、検査の性能を評価する。

Biomarker：バイオマーカー

通常の生物学的プロセス、発病過程又は治療的介入への薬理的応答の指標として、測定や評価される特性。

Classifier：分類指標

分類指標は、バイオマーカーの値を一連の予後分類に変換する数学的な関数であり、患者分類を可能にするマーカーとして定義することもできる。

Clinical Utility：臨床的有用性

マーカー又は検査が役立つことの臨床的有用性とは、

検査（マーカー）が介入の転帰の改善につながる可能性のことである。言い換えると、マーカー（GBM）が、他の標準的な臨床的特徴に加えて、マーカーで定義された集団の薬物反応や予後評価の予測を提供してくれる価値のこと。

Clinical Validity：臨床的妥当性

検査が臨床疾患や表現型の存在（又は不在）を予測する正確度を指す。

Genomic biomarker：ゲノムバイオマーカー

通常の生物学的プロセス、発病過程又は治療的介入への薬理的応答の指標として、測定可能な DNA 及び/又は RNA 特性。

(Biomarker) Qualification：(バイオマーカー) 適格性確認

適格性確認とは、提案されたバイオマーカーの用法の範囲内において、当該バイオマーカーが生物学的過程、反応あるいは事象を適切に反映し得ると判断され、医薬品又はバイオ技術開発におけるその使用が支持されるというバイオマーカーに関する評価結果を踏まえた結論。

Reproducibility：再現性

独立した従事者/研究者/研究室によって、正確に再現又は反復される検査又は実験の能力を指す。

