

用である。本稿では詳細を割愛するが、現在国内の6つの臨床試験グループ(JALSG, JCOG, J-CRSU, JGOG, JPLSG, WJOG)で、有害事象報告や監査の共通ガイドラインづくりを進めており、これらのガイドライン作成により国内グループの手順の標準化が進むことが期待される。

共同試験には単一のデータセンターにデータを収集する“single data center型”と、それぞれのグループのデータセンターにデータを収集し、後で統合解析を行う“multiple data center型”が存在する。前者のメリットはデータセンター間でデータ形式やデータマネジメントの手法などを擦り合わせる必要がないことであるが、自グループ外の参加施設の施設情報、研究者情報を管理する必要があり、それらの施設からの有害事象報告や監査についてもマネジメントする必要がある。これらのデメリットを考慮して、現在のところJCOGでは後者の“multiple data center型”を基本としているが、国際共同試験ではデータの質や施設マネジメントの観点から“single data center型”を求められる場合があり、“single data center型”への対応は今後の課題である。

JCOGでは、これまで韓国、米国と2試験の国際共同試験を実施しているが紙面の都合上、詳細は割愛する。国際共同試験とすることで確実にデータの質は落ちる上に、国別で治療効果に差がみられることはよく知られたデメリットである。にもかかわらず、個別化治療で細分化された治療開発を迅速に進めるには国際共同試験は避けては通れない課題であり、特に人種間差が少ないであろう韓国や台湾といった信頼できるアジア諸国との国際共同試験の体制構築を積極的に行っていく必要がある。

4. EDC(Electric Data Capture)システムの導入

これまでJCOGでは施設の研究者に対する負担増の懸念から紙のCRFでデータ収集を行ってきた。2012年2月時点で、JCOG参加554施設中、CRCがCRFを記入している施設は108施設(19%)にすぎず、いまだ8割の施設では医師自らがCRF記入を行っているのが現状である。

しかし一方で、JCOG全体で患者登録数およびそれに伴うCRFの枚数が飛躍的に増大しているという問題がある。実際2006年時点ではJCOG

全体の年間登録数は1,219例であったが、2011年時点での年間登録数は2,744例であった。また、追跡調査時に扱うCRFの枚数も2006年の20,775枚から2011年の48,325枚とこちらも急増している。今後も患者登録の増加が見込まれる中、人員増での対応も限界があり、データ収集と施設への問い合わせをより効率的に行うべく、2012年にEDCシステムの導入を決定した。2014年3月現在、患者登録のweb化はほぼ完了し、現在追跡調査のweb化および、一部試験での全CRFのweb化を進めているところである。

EDCシステムの導入には研究者の手間などのデメリットも存在するが、メリットとしてはEDC化によってこれまでリソースと優先順位づけの観点から実施できなかったComparative Effectiveness Researchの実施の可能性があげられる。たとえばJCOG試験へ登録されなかった患者についても最低限の背景情報と予後情報を収集し、臨床試験結果の一般化可能性を検討するといったタイプの一種のコホート研究についてEDCシステムを活用することが期待される。また、共同試験を行う際に“single data center型”のデータセンターを担うにはEDCシステムは必須であり、共同試験の観点からもEDCシステム導入のメリットはあると考えられる。

おわりに

本稿では詳細について触れることはできなかったが、このほかにも各グループにおける対象別の治療開発戦略を俯瞰する「治療開発マップ」の作成・公開や、既存のJCOG試験のデータを用いた「副次的解析」の推進など、JCOGでは研究の効率化や価値の最大化に資する取り組みを多面的に行っている。

これまでにあげてきた研究プロセスの効率化やアクティビティ向上のためのさまざまな取り組みは、米国ではNCI主導でトップダウンの形でさらに包括的かつ徹底的に実行されている。2015年4月に発足予定の「日本医療研究開発機構(いわゆる日本版NIH)」にその役割が期待されるが、現状NCIのような臨床試験グループを管理、支援する組織が日本では存在しない以上、各臨床試験グループが自助努力で研究プロセスの効率化、

アクティビティの向上に取り組む必要がある。さらに、そのノウハウを日本の臨床試験グループ間で情報共有することが重要であり、本稿がわが国の各臨床試験グループにおける「カイゼン」活動の一助となれば幸いである。

本稿の内容は、独立行政法人国立がん研究センターがん研究開発費 26-A-4(主任研究者：飛内賢正), 26-A-22(主任研究者：福田治彦)に基づく研究成果によるものである。

文 献

- 1) JCOG 基 本 規 約 [cited 2014 Apr 18]. Available from : URL : http://www.jcog.jp/basic/policy/A_020_0010_01.pdf.
- 2) 國枝太史, 中村健一, 柴田大朗, ほか. 日米欧のがん共同研究グループの試験における治療モダリティの比較 (JCOG1103-A). 日本癌治療学会 2011. 抄録番号 OS104-4.
- 3) 中村健一, 福田治彦. JCOG (日本臨床腫瘍研究グループ). 腫瘍内科 2010 ; 6 : 283.
- 4) Institute of Medicine. A National Cancer Clinical Trials System for the 21st Century : Reinvigorating the NCI Cooperative Group Program. 2010. Available from : URL : <http://iom.edu/Reports/2010/A-National-Cancer-Clinical-Trials-System-for-the-21st-Century-Reinvigorating-the-NCI-Cooperative.aspx>.
- 5) 木村 綾, 中村健一, 水澤純基, ほか. 多施設共同臨床試験におけるプロトコル作成期間の検討. 日本癌治療学会 2013. 抄録番号 109-4.
- 6) 厚生労働省. 臨床研究に関する倫理指針 [cited 2014 Apr 18]. Available from : URL : <http://www.mhlw.go.jp/general/seido/kousei/i-kenkyu/rinsyo/dl/shishin.pdf>.
- 7) 中村健一, 水澤純基, 阿部 純, ほか. 多施設共同臨床試験におけるセントラル IRB システムの導入と審査期間の検討. 日本癌治療学会 2013. 抄録番号 O109-3.
- 8) Go RS, Meyer CM, Mathiason MA, et al. Nature and outcome of clinical trials conducted by the Eastern Cooperative Group (ECOG) from 1977 to 2006 [abstract]. J Clin Oncol 2010 ; 28(15_suppl). Abstract No. 6069.
- 9) 片山 宏, 中村健一, 水澤純基. JCOG 論文迅速化プロジェクト. 日本癌治療学会 2011. 抄録番号 PS130-4.

* * *

特集

血液腫瘍に対する臨床試験のinfrastructure—現状の問題点と今後の方向性

がん多施設共同臨床試験 とバイオバンク：現状の 問題点と将来*

金戸 啓介**
高島 淳生**
中村 健一**
福田 治彦**
松村 保広***

Key Words : biobank, tumor repository, JCOG, clinical trial, cooperative group

はじめに

JCOGバイオバンク (Japan Clinical Oncology Group Biorepository) は、JCOG 試験に登録された患者を対象に、① 事前に計画された試料解析研究の有無によらず試料を収集し、② 試料を JCOG バイオバンクで一括保管し、③ 将来実施される試料解析研究に対して試料および診療情報を提供することを目的としたプロジェクトで、2013年7月より運用を開始した(図1)。

2014年3月末までにJCOGの11試験が参加し、約80名分の試料が登録された。

2014年度は約400～500名分の試料登録が見込まれる。

本稿では、JCOGバイオバンク設立の経緯や特徴、問題点と将来について述べる。

JCOGバイオバンク設立の経緯

JCOGでは、従来ホルマリン固定標本の未染色プレパラートを収集して行う免疫染色による研究や、凍結試料を用いた遺伝子解析などといった試料解析研究が実施されてきたものの、いずれも試

験単位で個別に計画され実施されたものであり、計画時の使用目的に限定されたものであった。

しかし、近年の分子生物学の発展はめざましく、試料解析研究計画時や実施中に見出されていなかった分子生物学的知見が、終了時には新たに判明していることも多く、臨床試験ごとに使用目的を限定した試料では、必ずしも十分価値の高い試料解析研究が行えなくなってきている。

さらに、試験ごとに試料を収集するには、そのつど実施体制の構築が必要となるが、試験ごとの体制構築は試料収集や保管の効率が悪くだけでなく、試料の収集・保管を担当する研究者の負担も大きい。

このような背景から、臨床試験ごとや専門領域別グループごとではなく、JCOG全体で共通の仕組みとしてのJCOGバイオバンクの必要性が1990年代後半から繰り返し議論されてきた。

2000年代になりJCOGバイオバンクの重要性の認識がさらに高まったため、国立がん研究センター内、JCOG試料解析研究委員会、JCOG病理委員会などで議論を重ね、2011年に国立がん研究センター東病院臨床開発センター内にJCOGバイオバンクを設置し、2013年7月に運用を開始した。

* Current status and future perspective of JCOG Tumor Repository.

** Keisuke KANATO, Ph.D., Atsuo TAKASHIMA, M.D., Kenichi NAKAMURA, M.D. & Haruhiko FUKUDA, M.D.: 国立がん研究センター多施設臨床試験支援センター/JCOG運営事務局・データセンター[〒104-0045 東京都中央区築地5-1-1]; Japan Clinical Oncology Group Operations Office, Multi-institutional Clinical Trial Support Center, National Cancer Center, Tokyo 104-0045, JAPAN

*** Yasuhiro MATSUMURA, M.D., Ph.D.: 国立がん研究センター東病院臨床開発センター新薬開発分野

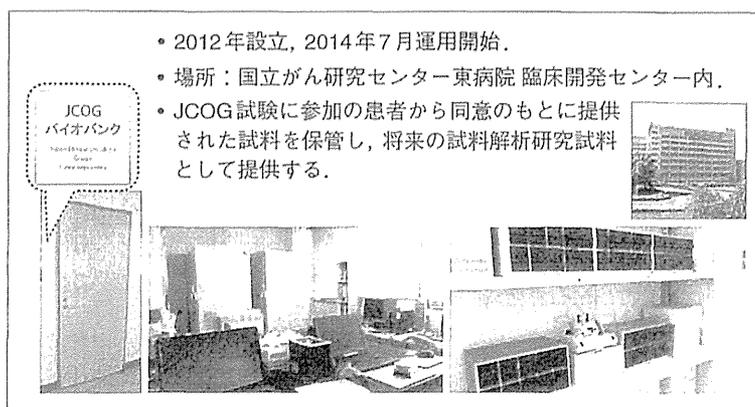


図1 JCOG バイオバンク

JCOG バイオバンクの特徴

1. 事前に計画された試料解析研究の有無によらずあらかじめ試料を収集する

JCOG 試験の患者登録中や患者登録終了以降に、重要な分子生物学的知見が明らかとなり試料解析研究が計画されることがあるが、試験開始後に新たに試料を収集しようとしても、試料の採取が不可能な場合や患者から同意を得ることが困難な場合が多い。

そうした状況を回避し、収集された試料から得られる知見の科学的価値を高めるためには、将来行われる試料解析研究の内容をあらかじめ限定しない、いわゆる包括的同意を患者から得た上で、試料提供を受ける仕組みが必要である。

JCOG バイオバンクでは、試験特有の事情により研究事務局/研究代表者やグループ代表者が希望しない場合を除いて、原則としてすべてのJCOG 試験を対象として、事前に計画された試料解析研究の有無によらず、試験開始とともに各患者からあらかじめ試料を収集する。

試験開始時点から試料収集を行うことにより、1試験あたりの収集試料の数を最大化して試料解析研究の精度が高められるだけでなく、原則すべてのJCOG 試験を対象とすることで、同一疾患を対象とした複数のJCOG 試験や、同じ薬剤を用いた複数のJCOG 試験の横断的な統合解析が可能となる。

2. 試料を一括保管する

多数の医療機関から試料収集を行おうとした

場合、各医療機関における試料処理過程や試料保存環境に施設間差があることが予想されるが、そうした試料の質的な相違は、将来実施される試料解析研究において正しい結果を得るのに障害となる可能性がある。

JCOG バイオバンクでは、試料の処理・搬送を一業者に一括委託するとともに、JCOG バイオバンクを国立がん研究センター東病院臨床開発センター内で一括管理し、試料の質的な相違の最小化と、全体的な作業の効率化を図っている。

3. 試料は質の高い診療情報と紐づいている

試料解析研究において正しい結果を得るためには、質的なバラツキが小さい試料とともに、質(精度)の高い診療情報が必要不可欠である。

JCOG 試験では、試験実施計画書の一定の適格規準を満たす患者のみを対象とし、規定に従った治療を行い、同じ規準を用いた有効性と安全性の評価が行われている。さらに、収集された診療情報はJCOG データセンターによりデータクリーニングが行われており、日常診療における診療情報に比べて、担当医の医学的意思決定や評価の偏りに起因するデータの系統的なバイアスの小さい診療情報を有している。

JCOG バイオバンクの概要

JCOG バイオバンクの概要を以下1~5に示す(図2)。

JCOG バイオバンクでは、試料登録、抽出、バイオバンクへの入庫、試料解析研究機関への出庫といった一連の流れをJCOG Web Entry System

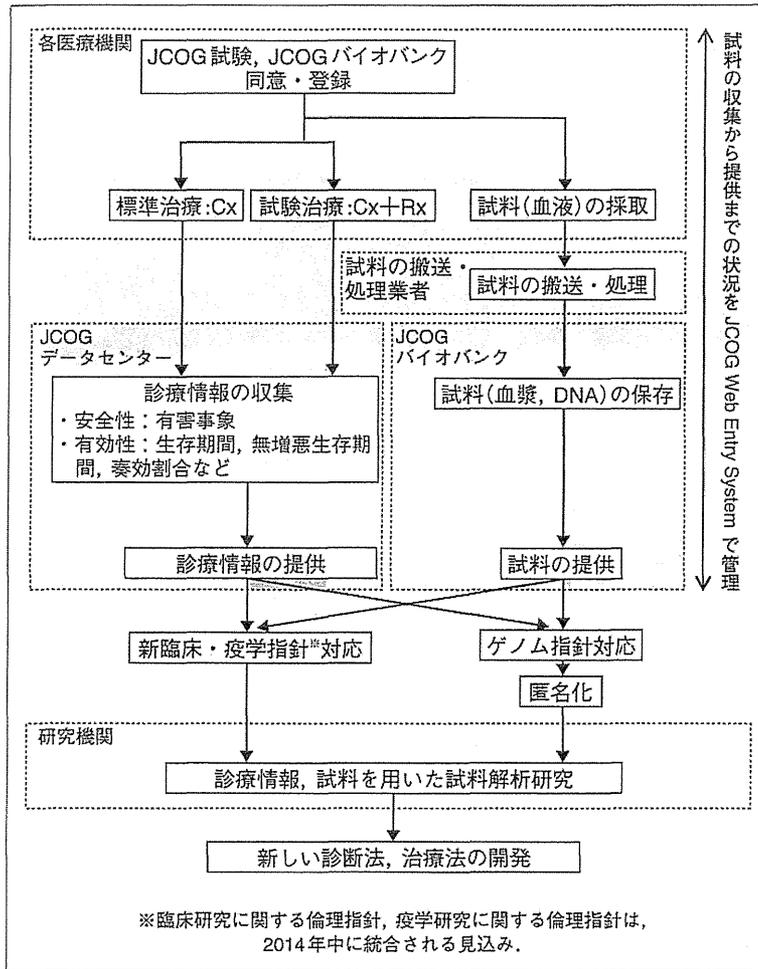


図2 JCOG バイオバンク試料収集から将来利用されるまでの流れ

で一括管理している。これは、米国のがん臨床試験グループである Southwest Oncology Group (SWOG) の Specimen Tracking System をモデルとした¹⁾。

1. 対象

JCOG 試験に登録された患者のうち、JCOG バイオバンクへの試料の提供と将来の試料解析研究での利用について同意が得られた患者を対象としている。

2. 試料

(1) 血液(全血) 14 ml

(2) 日常診療における保存病理組織

保存病理組織の収集については、現在、同意を取得するのみとしている。

収集形態、収集のタイミング、各医療機関への

インセンティブ等詳細を詰めた上で運用を開始する(現在調整中)。

3. 方法(試料の収集・保管・提供)

(1) 試料の収集

① 試料の登録

患者からの同意取得後、担当医は、JCOG web tracking system で患者登録に続いて試料登録、および送付用の試料収集伝票をダウンロードする。

② 試料の採取、提出

血液(全血)：あらかじめ医療機関に配布した専用採血管を用いて血液 14 ml を採取する。採取後、試料収集伝票とともに試料搬送・処理業者へ提出する。

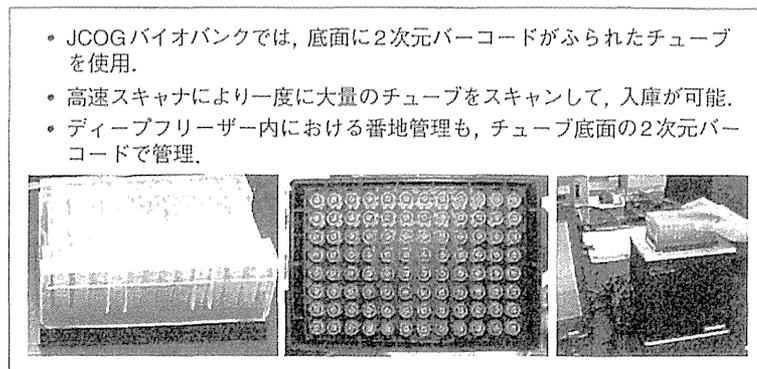


図3 JCOGバイオバンクでの試料入庫

③ 試料の処理

試料搬送・処理業者は、血漿分離を行い血球成分からDNAを抽出し、血漿およびDNAを専用チューブに分注する。

専用チューブは、JCOGバイオバンク番号が付された2次元バーコードラベルで管理する。

JCOGバイオバンク番号は、各試料識別のため、JCOG試験番号、JCOG患者登録番号、試料の種類、採取時期、試料のシリアル番号を組み合わせた番号である。

(2) 試料の保管

JCOGバイオバンクでは、試料をディープフリーザー(−80℃)で保管する。試料は、JCOGバイオバンク番号、および専用チューブ底面に付された2次元バーコードで識別、管理する(図3)。

(3) 試料の提供(予定)

JCOG内での審査機構および関連施設のIRBで承認された試料解析研究に対して、バイオバンクで保管している試料を提供する。

① 試料解析研究が「ヒトゲノム・遺伝子解析研究に関する倫理指針」の適用となる場合

JCOGバイオバンクは試料を、JCOGデータセンターは必要な診療情報を、JCOGバイオバンク個人情報管理者(国立がん研究センター東病院院長)へ提出する。JCOGバイオバンク個人情報管理者は、JCOGバイオバンク番号の連結可能匿名化を行う(JCOGバイオバンク番号からJCOGバイオバンク匿名化番号へ置き換える)。JCOGバイオバンク個人情報管理者は匿名化された試料および診療情報を試料解析研究実施機関へ提供する。

② 試料解析研究が「疫学研究に関する倫理指針」の適用となる場合

JCOGバイオバンクは、試料解析研究実施施設へ必要な試料を提供する。試料の識別にはJCOG登録番号を用いる。JCOGデータセンターは、必要な診療情報を試料解析研究実施施設に提供する。患者の識別番号としてはJCOG登録番号を用いる。

4. 試料の保管期間

試料は、試料解析研究により使い切られるまで保管される予定であり、保管期間の期限は定めていない。

5. 参加研究機関

JCOG試験参加施設のうち、JCOGバイオバンクへの参加に関して医療機関の承認が得られた施設が参加している。

JCOGバイオバンク現状の問題点と将来

1. 現状の問題点

JCOGバイオバンクの試料は、質の高い診療情報を有していることから、それらと分子生物学的異常との真の関連を正しく検出できる可能性が高く、将来JCOGバイオバンクの試料を用いた試料解析研究の成果が期待される。

しかし、より試料価値を高め、かつより円滑な運用を行うためには、(1)～(3)にあげるような課題の解決が必要である。

(1) 試料の種類、採取回数の拡充、インフラ整備

2013年7月より運用を開始したJCOGバイオバンクは、全JCOG試験が参加しやすくすること

を優先し、実施計画書では、

- ・血液のみを採取(血漿, DNAのみを収集)
- ・採血回数: 1ポイント(原則として治療開始前)のみ

と規定している。しかし、これらの試料のみでは将来実施できる試料解析研究に限られるため、JCOGの研究者からは以下のような要望が出されている。

- ・採血回数を複数回可能としてほしい。
- ・血液から抽出したDNAのみでなく、RNAもバイオバンク試料としてほしい。
- ・保存病理組織(パラフィンブロック)や凍結組織もバイオバンク試料としてほしい。

これらをバイオバンク試料として収集することは重要だが、恒常的に収集、保管するためのバイオバンクを構築し運用するためには、JCOG内部のみならず各医療機関内での調整(液体窒素タンクの設置といったインフラ整備、臨床医や病理医へのインセンティブなど)が必要である。

また、現行の血液試料(血漿, DNA)の保管もディープフリーザー2機で運用しており、近い将来、増設が必要である。

(2) 人員の確保

現在JCOGバイオバンクは、以下の人員構成で運用している。

- ・バイオバンクでの試料保管: 2名[国立がん研究センター東病院]。
- ・バイオバンクの運用(実施計画書/説明文書・同意書の作成, 医療機関の問い合わせ窓口, 試料処理搬送業者との調整, Web Entry Systemのメンテナンスなど): 3名[国立がん研究センター中央病院, JCOGデータセンター/運営事務局]。

マンパワーが非常に限られており、「(1)試料の種類, 採取回数の拡充, インフラ整備」を拡充していくためには、さらなる人員の確保も必要である。

(3) 運用資金の確保

JCOGバイオバンクは2014年4月現在、独立行政法人国立がん研究センターがん研究開発費「医療イノベーション推進を目指した国立がん研究センター等バイオバンクの設立と運用」[主任研究者: 金井弥栄(研究所副所長/分子病理分野分野

長)]の一部で運用しているが、インフラ整備や研究者へのインセンティブ、マンパワー等を確保するには十分ではない。

2. 将来—JCOGバイオバンクの将来展望—

JCOGが実施している臨床試験は、約6割がランダム化第III相試験であり、これらの試験で得られた診療情報と紐づいたJCOGバイオバンク試料を用いた試料解析研究では、単なる予後因子だけでなく、治療効果予測因子の探索も可能となる。

さらにJCOG試験では、試験実施計画書で規定された、均一性の高い集団、一定の治療方針、一定の規準による評価(有効性・安全性)が行われており、得られた診療情報もJCOGデータセンターでクリーニングされた質の高いデータである。

JCOGバイオバンク試料を用いた解析研究により、将来以下のような成果(可能性)が期待される。

- ・採血で得られたDNAや血漿タンパクにおける分子レベルの変化と、JCOG試験における毒性データとの関連を検討することで、毒性発現に関する予測因子の探索や検証が行える。
- ・毒性発現の予測因子が見出された場合、特定の患者が過度の毒性を回避し、より適切な治療を受ける治療の個別化につながる可能性がある。
- ・腫瘍組織における遺伝子やタンパク質などの分子レベルの異常と、JCOG試験によって得られた腫瘍縮小効果等の有効性データとの関連を検討することで、腫瘍の増殖や増殖抑制に関係している新たな分子標的の探索や、既知の分子標的の役割の検証が行える。
- ・新たな分子標的が見出された場合には、分子標的薬の創薬につながる可能性がある。
- ・腫瘍組織における分子レベルの異常と有効性データとの関連(交互作用)を治療別に検討することで、新たな治療効果予測因子(バイオマーカー)の探索や、既知の予測因子の検証が行える。
- ・新たなバイオマーカーが見出された場合には、バイオマーカーを用いた治療の個別化につながる可能性がある。
- ・腫瘍組織における分子レベルの異常と予後と

の関連を検討することで、新たな予後因子の探索や、既知の予後因子の検証が行える。新たな予後因子が見出された場合には、治療選択における医師や患者の意思決定に寄与できる可能性がある。

しかし、これら可能性をより確実なものとするためには、「1. 現状の問題点」に列記した課題を解決しなければならず、そのためには、何より

もまずJCOGバイオバンク運用のための恒常的な研究費確保が必要である。

文 献

- 1) Hoban CJ, Franklin W, Kopeccky KJ, Baker LH. SWOG Cooperative Group Biorepository Resource : Access for Scientific Research Studies. Clin Cancer Res 2011 ; 17 : 5239.

* * *

ベネフィット・リスク評価の実際 —研究者主導のがん臨床試験の事例紹介—

中島 章博^{*1}, 小森 哲志^{*1}, 神浦 俊文^{*1}, 杉谷 康雄^{*1}, 木村 友美^{*1}, 福田 治彦^{*2}

Benefit Risk Assessment

An Example Using Results from an Investigator-initiated Clinical Trial

Akihiro NAKAJIMA^{*1}, Tetsushi KOMORI^{*1}, Toshifumi KAMIURA^{*1}, Yasuo SUGITANI^{*1},
Tomomi KIMURA^{*1} and Haruhiko FUKUDA^{*2}

本稿は、2014年1月14日に医薬品医療機器レギュラトリーサイエンス財団主催の第149回レギュラトリーサイエンス エキスパート研修会『ベネフィット・リスク評価の実際』で、実例を用いて参加者とともに行ったワークショップの様子をまとめたものである。

研修会当日は、ベネフィット・リスク評価の基本となるフレームワーク¹⁾²⁾とその使い方、及び Multiple Criteria Decision Analysis (MCDA) に代表される定量的手法の基本的考え方を説明した。その後、がんの医師主導研究の二つの実例をもとに、まず演者が行ったベネフィット・リスク評価結果を一例として紹介した。その後、2-3人で1組となった参加者が、どちらかの例を選んで評価のプロセスを体験するワークショップを行った。また、限られた時間で項目の取舍選択やスコアリング、重みづけなどを体験してもらうために、あらかじめ Microsoft Excel[®] を使って作成した視覚化ツールを用意し、データも入力したうえで提供した。

以下では当日紹介した二つの試験のうち、進展型小細胞肺癌の治療法を比較した JCOG (Japan Clinical Oncology Group: 日本臨床腫瘍研究グループ) 9511 試験でのベネフィット・リスク評価の事例を紹介する³⁾。

なお、医薬品のベネフィット・リスク評価の基本については『ベネフィット・リスク評価入門』⁴⁾、定量的手法である MCDA については『ベネフィット・リスク評価 中級編』⁵⁾

(いずれも日本製薬工業協会ウェブサイトにて無料で公開中)をご参照いただければ幸いである。

1. 事例の紹介

JCOG9511 試験は、進展型小細胞肺癌に対するエトポシド+シスプラチン (EP) 療法とイリノテカン+シスプラチン (IP) 療法の多施設共同ランダム化第 III 相比較試験である。予定登録数を 230 例として 1995 年に開始したが、中間解析の結果、多重性を考慮しても統計学的に有意な差が生存期間に認められたため、登録数 154 例で早期中止され、IP 療法が進展型小細胞肺癌に対する新たな標準治療であると結論付けたものである。生存期間の中央値は IP 群が 12.8 か月、EP 群が 9.4 か月であった ($p=0.002$: 多重性調整しないログランク検定)。重篤な骨髄抑制は IP 群よりも EP 群で頻度が高く、重篤な下痢は EP 群よりも IP 群で頻度が高かった。

以下、代表的なベネフィット・リスク評価のフレームワークに沿って演者側から提示した内容を説明する。フレームワークは考える道筋や順番を提案するガイダンスで、いくつかのステップからなる。すなわち、①背景の定義、②ベネフィットやリスクの項目の列挙、③データソースの特定、④項目の選択、重みづけ、⑤結果のまとめと意思決定である。

^{*1} 日本製薬工業協会臨床評価委員会データ・サイエンス部会タスクフォース 1 東京都中央区日本橋本町 3-4-1 (〒103-0023)

^{*2} 国立がん研究センター多施設臨床試験支援センター/JCOG データセンター 東京都中央区築地 5-1-1 (〒104-0045)

1.1 背景の定義

ベネフィット・リスク評価の結果は、その目的や意思決定者など背景情報によって異なり得る。すなわち疾患の重篤度やアンメット・ニーズによって、同じデータでもその医学的価値は異なる可能性がある。したがって、疾患や医薬品の情報など、意思決定の背景情報を定義することが極めて重要である。Table 1にJCOG9511試験のデータをもとにした意思決定の背景情報をまとめた。

当日のワークショップでは、評価する立場をスポンサー（例えば企業など）だけでなく、規制当局、患者や医療従事者など自由に選んで評価を行ってもらった。

1.2 データソースの特定、ベネフィットやリスクの項目の列挙と選択

通常なら、ここで複数の臨床試験からいくつかを選んだり、あるいはその全てを利用したりするが、今回はJCOG9511試験の情報のみを参考にした。ここで、ベネフィットやリスクの項目を列挙し、選択した。BRAT (Benefit Risk Action Team) はデータソースを特定する

際、試験デザインの特徴、試験の質の尺度（例えば、Jadad Scale）などの情報も含めたデータソーステーブルを作成することを提案している⁹⁾。意思決定に影響を与える項目として、ベネフィットは生存期間、リスクは治療関連死、米国 National Cancer Institute (NCI) による共通毒性規準 (Common Terminology Criteria for Adverse Events (CTCAE)) での Grade 4 の非血液毒性 (生命を脅かす有害事象) などを挙げ、バリュー・ツリーを作成した (Fig. 1)。バリュー・ツリーとは、項目をベネフィットとリスクにわけて階層構造状のツリーで表示したものである。

ワークショップでは、付箋紙にベネフィットやリスクの項目を一つずつ記載し、それらに優先順位をつけて並べ替えたり、取捨選択したりしながら、それぞれのグループのバリュー・ツリーを作成してもらった。今回はJCOG9511の試験結果を提示した上で、項目の列挙と選択を同時に行ったが、データソースを特定する前に項目を列挙したほうが、データに影響されることなく、重要な項目を列挙できるだろう⁹⁾。

Table 1 JCOG9511 試験の背景情報

ベネフィット・リスク評価を行う立場：	試験実施者である JCOG (スポンサー)
意思決定の目的 (=JCOG9511 の試験目的)：	進展型小細胞肺癌に対するよりよい治療法の選択 <ul style="list-style-type: none"> • エトポシド+シスプラチン (標準治療) • イリノテカン+シスプラチン (新治療)
対象疾患： 進展型小細胞肺癌	重篤度 (当時の) 標準治療の効果 罹患率
	増殖速度が速く、予後が不良 <ul style="list-style-type: none"> • 3年生存割合：10%程度 • 奏効割合：75%程度 • 完全奏効割合：10%程度 肺癌の年間発症数は男性で約 65,000 人、女性で約 30,000 人。 そのうち、小細胞癌は 10~15%。
	アンメット・ メディカルニーズ
	<ul style="list-style-type: none"> • 生存時間の延長 • 重篤な有害事象の軽減 • 有効な分子標的薬なし
	対象集団
	<ul style="list-style-type: none"> • 組織診又は細胞診で小細胞肺癌の診断 • 進展型の診断 • 前治療なし • 測定可能又は評価可能病変を有する • PS*：0, 1, 2 • 少なくとも 3 か月以上の生存が可能と考えられる • 年齢：70 歳以下 • 骨髄、肝、腎機能が保持されている • 重篤な心肺合併症、3 年以内に活動性の重複癌がない
医薬品の情報	用法・用量
	<ul style="list-style-type: none"> • EP (エトポシド+シスプラチン) <ul style="list-style-type: none"> ➢ エトポシド 100 mg/m² day1,2,3 ➢ シスプラチン 80 mg/m² day1 ➢ 3 週ごとに 4 コース • IP (イリノテカン+シスプラチン) <ul style="list-style-type: none"> ➢ イリノテカン 60 mg/m² day1,8,15 ➢ シスプラチン 60 mg/m² day1 ➢ 4 週ごとに 4 コース

*Eastern Cooperative Oncology Group (ECOG) performance status

JCOG9511の Value Tree

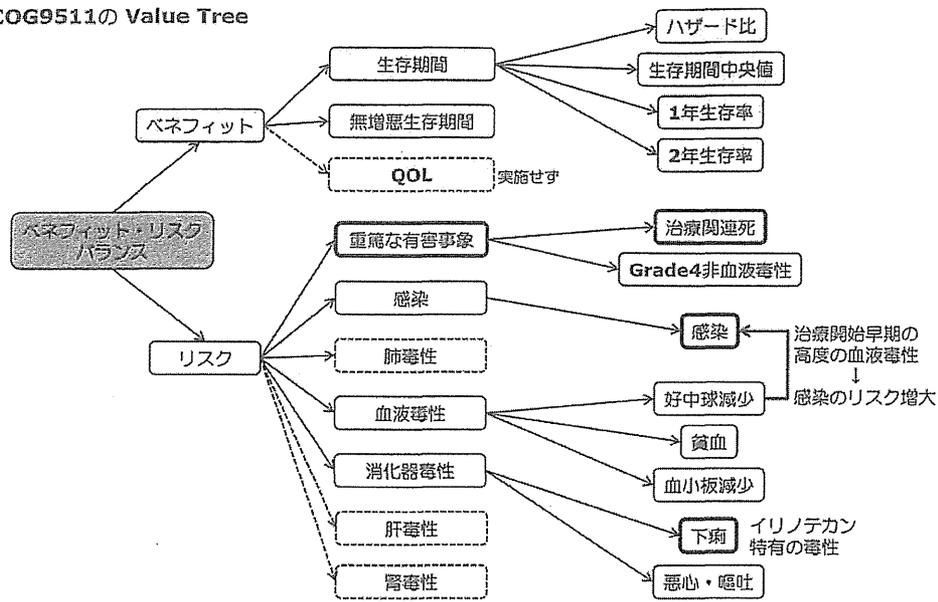


Fig.1 バリュー・ツリーの例

1.3 重みづけと結果のまとめ(視覚化)

演者からは定量的手法を用いた重みづけは提示せず、定性的に優先順位をつけた例を提示した。最後に結果をまとめ、半定量的手法⁹⁾を用いてフォレスト・プロットによる視覚化を行った例を示した(Fig. 2)。各項目のリスク差の点推定値と95%信頼区間をプロットし、上に位置する項目ほど優先順位が高いと考えた。

なお、臨床的にはハザード比が生存時間の評価指標とし

てよく使用されるが、今回は2年生存率及び1年生存率の差を用いて示している。演者の解釈として、IP療法により享受できる生存期間延長(ベネフィット)は、治療関連死とGrade4非血液毒性という重篤な有害事象のリスク増加よりも大きいと考えること、下痢、感染は注意深く観察し、支持療法の充実を今後図る必要があること(註：その後支持療法の進歩により、下痢のコントロールが可能となった)などを紹介した。

JCOG9511の Forest Plot

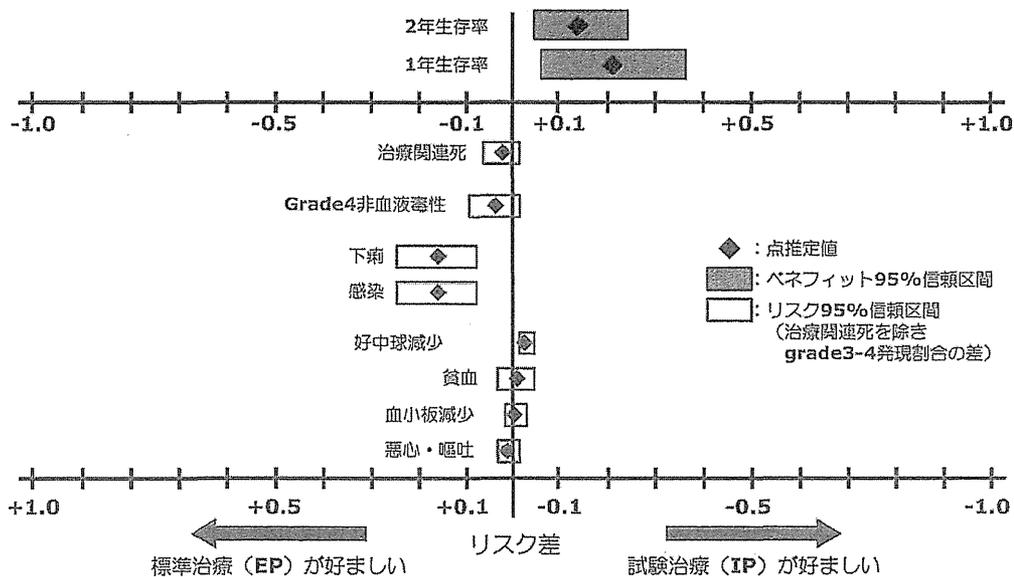


Fig.2 フォレスト・プロットの例

2. 定量的手法体験ワークショップ

以上のように、演者側からは意思決定の背景、試験の概要及び半定量的(定性的)方法を用いて行ったベネフィット・リスク評価の事例を紹介した。実際はこれで十分な場合が少なくない⁶⁾。しかし、このワークショップではMCDAに代表される定量的手法を試してみたい参加者のために、簡単な説明とツールを提供した。MCDAの基本的な説明はタスクフォースの報告書⁷⁾に譲るが、希望者には以下の三つの手順で考えてもらった。

2.1 スコア化

1.2項で挙げた個々の項目に対し、その項目が取る値の大きさを医療上の価値の大きさに翻訳する作業である。例えば、ベネフィットの項目の一つである2年生存率が0%から30%の範囲で変化し得ると考え、生存率ごとの価値の大きさを評価する。2年生存率が0%とき価値は最小で0点、また30%のとき価値は最大で100点等と定めよう。Fig. 3 Aの左側のグラフでは、2年生存率が大きいほど比例的に価値も大きいと評価している。一方、右側のグラフでは、2年生存率が10%より小さいか大きいかで、価値が増加する程度(直線の傾き)が異なると評価している。

同様に、例えばリスクの項目の一つである治療関連死の発現割合について、やはり発現割合の値と医療上の価値との関係を検討する。Fig. 3 Bのグラフのように、発現割合が大きいほど治療上の価値は比例的に小さくなる、というのも一つの評価である。また、Fig. 3 Bの右のグラフのよ

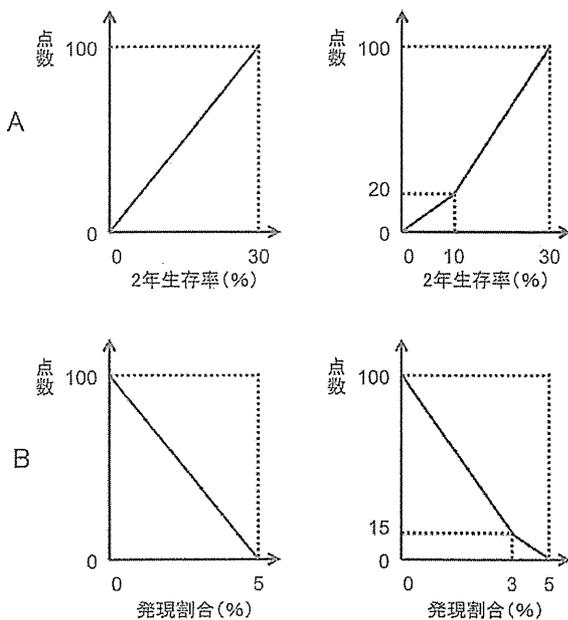


Fig. 3 2年生存率(A)及び治療関連死(B)のスコア化の例

うに、発現割合が3%より小さいか大きいかで比較して、医療上の価値の減少の程度が異なると判断する場合もあるかもしれない。

このように、スコア化を通じて、個々の項目を自分ならどのように医療上の価値と結び付けているかを明確にする。

2.2 重みづけ

データと医療上の価値との関係を大まかに決めたら、次に複数の項目間の相対的な重要性の違いを0から1までの値を取る「重み」として定める。今回は、全ての項目の重みの合計が1となるように定めたツールを配布した。先の例では、2年生存率については0%の場合を0点、30%の場合を100点と定め、また治療関連死については5%の場合を0点、0%の場合を100点と定めていた。

すなわち、2年生存率については生存者が30%増えることが100点の差として表されており、また治療関連死については死亡者が5%減ること(あるいは生存者が5%増えること)が100点の差として表されている。どちらも100点の違いではあるが、これらが医療上等価であるとは考えにくい。両者を天秤にかけたときに、どちらがどれくらい重いかという判断を、それぞれの項目対ごとに考えてもらった。これは、ある項目における100点という価値を、別の項目における100点と相対比較する作業であり、一方が他方に対して何倍価値があるか、というトレード・オフの関係を定めることでもある。中には、ある項目の相対的重要性を減らしたところ、結果が思わぬ方向に変わったので驚いたという参加者がいた。その理由を探ってみたところ、その項目の重みを小さくした分、他の項目の重みが相対的に大きくなったためであった。

このように、重みの評価は、複数の項目を同時に考慮しながら絶対的な評価を定める作業であり、こうして定めた重みを通じて、その評価者がどの項目をどの程度重視しているかを明確にする。

2.3 合計スコア

ワークショップで配布したツールは、項目ごとの「スコア×重み」を全て足し合わせて合計スコアを算出するものであった。また、全ての項目の合計スコアだけでなく、ベネフィットとリスク、それぞれの総合スコアを算出することもできた。このような計算結果の例をFig. 4及びTable 2に示した。

Best, Worstは変化しうる範囲の最良値(価値が最大となる値)と最悪値(価値が最小となる値)であり、項目ごとにこの範囲でスコア化を行う。項目間の相対的な重要度(重み)を検討し、Table 2のようなEffect Tableにまとめ、

Table 2 総合スコア算出

項目	変化しうる範囲		重み	試験結果		単位
	Best	Worst		イリノテカン +シスプラチン(IP)	エトポシド +シスプラチン(EP)	
ベネフィット	合計					
生存期間						
中央値	15	8	0	12.8	9.4	month
1年生存率	70	20	100	58.4	37.7	%
2年生存率	40	0	100	19.5	5.2	%
無増悪生存期間						
中央値	8	3	0	6.9	4.8	month
6か月無増悪生存率	80	20	0	65.3	35.6	%
12か月無増悪生存率	15	5	0	12.5	7.9	%
リスク	合計					
重篤な有害事象						
治療関連死	0	5	30	4.0	1.3	%
Grade4 非血液毒性	0	15	10	6.7	2.6	%
感染						
感染	0	30	5	5.3	3.9	%
血液毒性						
好中球減少	0	100	2	65.3	92.2	%
貧血	0	50	1	26.7	29.9	%
血小板減少	0	30	1	5.3	18.2	%
消化管毒性						
下痢	0	30	5	16.0	0.0	%
悪心・嘔吐	0	30	1	13.3	6.5	%

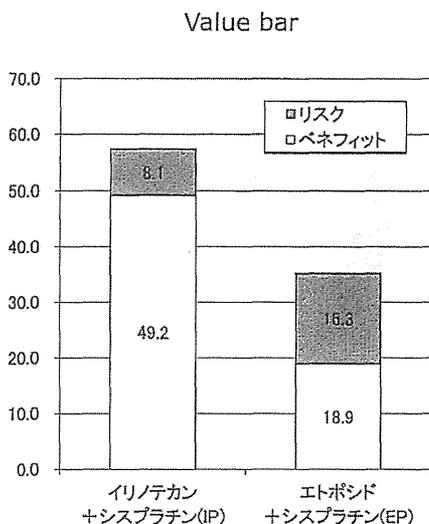


Fig. 4 総合スコア視覚化の例

合計スコアを算出する、合計スコアは大きいほど価値が高く、Fig. 4のような Value bar を用いて視覚化する。

3. 感度分析

ワークショップは2人(あるいは3人)1組で行ってもらったが、その中でも各項目の相対的重要性や解釈に違い

が見られた。当日はツールを提供したので、それぞれの組内で意見が割れた場合には、両方の意見を試してみることも(=感度分析)も体験してもらった。演者側でもあらかじめスコアリングや重みづけに以下のようないくつかのバリエーションを仮定して、それぞれがベネフィット・リスク評価で得られた結論にどのような影響を与えるのかを確認する感度分析を実施し、ワークショップの後に紹介した。

3.1 重要な項目の重みを0~100%まで変化させる

リスクの項目のうち、治療関連死を特に重要な項目として重みを変化させ、結果の変化を Fig. 5 a に示した。治療関連死の重みが37%以上となった場合に結果が逆転し、EP群のベネフィット・リスク・バランスの方が優れている結果となった。とは言え、あらためて治療関連死の重みを37%以上とすることの意味を考えてみると、これは、例えば治療関連死が5%増えることを、2年生存率や1年生存率が50%~70%程度減少することと同程度と評価することに相当する。このような評価が妥当と考えられるケースは少ないと考えられるため、結論が逆転することはなさそうだ、と言えそうな結果である。

3.2 重要な項目の結果の値を、信頼区間の下限から上限まで変化させる

今度はベネフィットの項目のうち、各群の2年生存率の

値を信頼区間下限から上限まで変化させた場合の結果の変化を Fig. 5 b に示した。変化させた範囲内のどのような値をとった場合でも、結果が逆転することはない。

3.3 バリュー・ツリーに一項目追加する

重要性や一貫性の観点からバリュー・ツリーから落としていた項目があった場合に、それらを追加することが結果

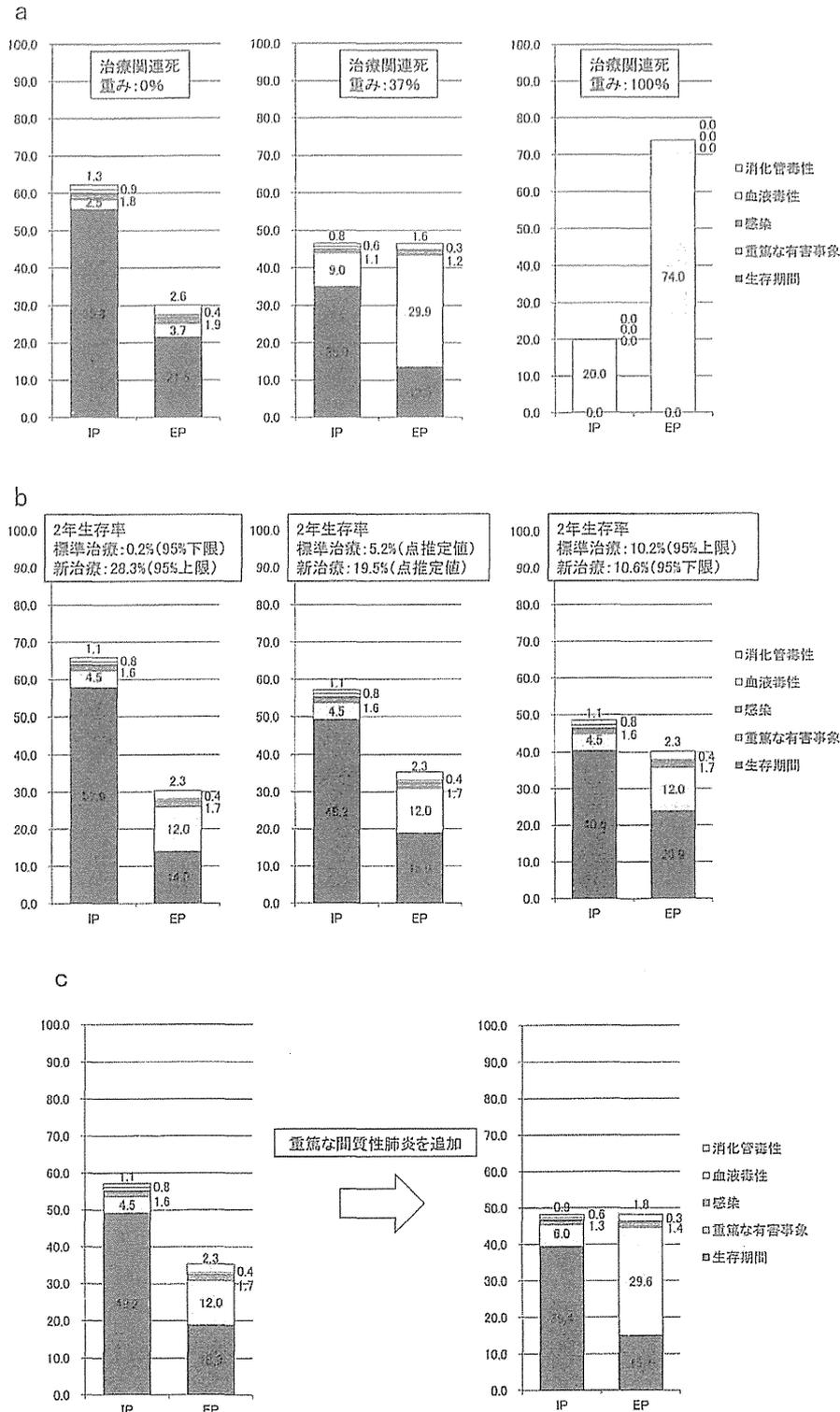


Fig. 5 感度分析の事例

に影響を与えるかを検討するものであり、今回は結果に影響を与えるものはなかったが、仮想的な例として「重篤な間質性肺炎」がIP群で4.4%発現し、EP群では0%であったという情報があり、Bestを0%、Worstを5%としてスコア化し、重みを20%とした場合、結果がどのように変化するかをFig. 5cに示した。重篤な間質性肺炎を追加した結果、IP群とEP群でベネフィット・リスク・バランスがほぼ同じになった。ベネフィット・リスク評価にこれほど大きな影響を与える項目がもし漏れていたなら、こういった項目はバリュー・ツリーに追加すべきである。

ここで取り上げた例以外にも、ベネフィット全体の重みを0～100%に変化させたり、スコア化に影響するBest・Worstの値を変化させる感度分析などもある。特に、不確かさが大きい要素について感度分析を行う必要があるだろう。

4. 考察

ベネフィット・リスク評価は新しく古い話題である。最近欧米の複数のグループから、医薬品のベネフィット・リスク評価に応用可能なフレームワークが相次いで提案されたことから、それぞれの医薬品に関して具体的な取り組みが行いやすくなった背景がある。今回のワークショップではその中の一つ、米国PhRMAから提案されたBRATのフレームワークを主に参考にした定性的(半定量的)な手法を用いて、日本のがんの臨床試験の実例を用いたベネフィット・リスク評価の事例を紹介し、更にその評価をもとにMCDAを中心とした定量的手法にも触れてもらった。

今回のワークショップでは、がん領域の事例を紹介したが、これにはいくつかの理由がある。まず、がん領域ではベネフィットやリスクの項目が国際的に標準化されており、複数の試験間の比較が行いやすい。例えばベネフィットとしては通常生存期間の延長、リスクとしてはNCI共通毒性規準を用いた評価項目が、国や試験によらず広く用いられている。ただ、通常ベネフィットの指標として用いるハザード比をリスクと同じスケールのフォレスト・プロットで表現しにくいなどの限界点もある。更に対処可能な有害事象やリスク最小化策の効果が期待できる安全性懸念事項に関しては、利用するデータの幅やその重みづけに工夫が必要である。また、フォレスト・プロット上で表現されないアンメット・メディカルニーズなどの重要な背景情報についても、総合的に判断して最終的な評価を行う必要がある。

なお、当時はQoLの評価項目を取得していなかったが、当日のワークショップではこの評価項目を重要視してベネ

フィット・リスク評価を行った参加者が多かった。

MCDAに関しては、上述の通り必須ではないし、かなりの人手や時間がかかる。更に、定量的な方法を用いることでより客観的な評価ができるはずだ、などと考えるのは短絡的である。むしろ、本来主観的に行うほかないベネフィット・リスク評価の中身を、可能な範囲で、あるいは最小限必要なところだけ定量化して見やすく、もしくは「透明に」している、と考えた方が実情に近い。もし行うとすれば、例えば同時に複数の項目を評価する場合、突出して「明らかに重要」といえるような項目がなく項目間の重要性が競合する場合、議論を重ねても評価者によって判断が異なる場合、不足情報等がありデータの確からしさが疑わしい場合などに効果があるかもしれない。

また、スコアリングのモデルなどの「正しさ」に不安が残る場合もあるかもしれないが、そもそも「正しい」関数を完璧につくることが本来の目的ではない。逆にこうした「関数」を申請時に作成しておけば、市販前から市販後まで通してベネフィット・リスク評価の履歴管理ができる。有効性・安全性の情報を更新する必要が生じたら、関数を入力していた値を再検討・変更したり、ある評価項目の医学的な重要性を見直す必要が生じたら、関連する評価項目の重みを再検討・変更したり、未知の重篤な有害事象の情報が得られたら、評価項目の一つ追加して(スコア化等も行い)関連する評価項目の重みも再検討・変更したりすればよい。透明性を高めるために、審査時に使った関数を公開し、市販後のベネフィット・リスク再評価でも運用することも考えられる。更に、3章で述べた感度分析は現状の結果について実施したものであったが、近い将来に新しいデータが得られた場合に、どのようにベネフィット・リスク評価の結果が変わるかを感度分析によって検討することも可能である。例えば重要な有害事象が新たに見つかった場合に、どのように変化するかを、3.3項と同様の方法で検討が可能であるし、もし医療環境の進歩により下痢がコントロール可能になったら下痢の重みももっと低くなるかもしれないが、その場合に評価結果がどのように変化するかを、3.1項と同様の方法で検討できる。こういった検討を行うことで、新しい情報が得られる前に先手を打った対応がとれる場合もあるかもしれない。

当日のワークショップには80名近くの参加者がおり、全ての組で行われた評価を把握することはできなかったが、それでもいろいろな意見が出たことはうかがえた。もっと大人数で行った時には、更に多くの、そして大きな差異がみられることだろう。したがって定性か定量かによらず、感度分析を常に検討すべきである。

3章で紹介した以外にも様々な感度分析の方法があり、例えば別の立場の方々にベネフィット・リスク評価を実施

してもらったり、今回のワークショップで体験してもらったように、検討した項目の取捨選択や相対的重要性が変わるとどのように結果が変わるのか、なども感度分析の一種である。様々な感度分析を行うことで、自身が実施したベネフィット・リスク評価の意味をより深く理解することに大きな意味がある。

MCDaを用いた定量的手法では、実際には価値判断がより細かく繰り返し実施される分、感度分析の重要性が更になる。Hiviewなどのツールを用いることで、重みづけの調整などが評価の結論にどのように影響するかを視覚化することもできる。こういった様々な視覚化ツールを使うことは感度分析の負担を下げ、ステークホルダー間の相互方向の合意形成に役立つだろう。

おわりに

医薬品のベネフィット・リスク評価は2014年3月現在、新たなICHのテーマ候補としても提案されており、今後当局や各企業、又はアカデミアなどで継続して取り組まれていく課題であると考えられる。定量的手法を用いた解析を行う場合でも、患者や医療従事者の定性的(質的)な評価が各項目のスコアリングや重みづけに必須であることから、事前の選考に関する調査を積極的に行う必要がある。また、がん領域のように比較的国際的標準化が進んだ領域ばかりではないことも考慮する必要がある。

このように臨床試験計画の段階から申請時のベネフィット・リスク評価を念頭に置いて計画することが重要で、ここで紹介した事例がその一助となれば幸いである。

文 献

- 1) Coplan, P.M., *et al.*, Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines. *Clin Pharmacol Ther.* 2011, 89 (2), p. 312-315.
- 2) Liberti, L., McAuslane, N., Walker, S. Progress on the development of a benefit/risk framework for evaluating medicines. *Regulatory Focus.* 2010, p. 1-6.
- 3) European Medicines Agency, Benefit-risk methodology project. Work package 4 report: Benefit-risk tools and processes. [http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/03/WC500123819.pdf]. May 9 2012.
- 4) FDA, Structured approach to benefit risk assessment in drug regulatory decision making. Draft PDUFA V Implementation Plan. <http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM329758.pdf>, February 2013.
- 5) Noda, K., *et al.*, Irinotecan plus cisplatin compared with etoposide plus cisplatin for extensive small-cell lung cancer. *N Engl J Med.* 2002, 346 (2), p. 85-91.
- 6) 日本製薬工業協会データサイエンス部会, ベネフィット・リスク評価 入門. 2013. <http://www.jpma.or.jp/about/board/evaluation/allotment/pdf/benefitrisk.pdf>.
- 7) 日本製薬工業協会データサイエンス部会, ベネフィット・リスク評価 中級編 多基準決定分析への招待—その理論と事例—. 2013. http://www.jpma.or.jp/information/evaluation/allotment/pdf/benefitrisk_mid.pdf.
- 8) Levitan, B.S., *et al.*, Application of the BRAT framework to case studies: observations and insights. *Clin Pharmacol Ther.* 2011, 89 (2), p. 217-224.
- 9) Levitan, B., Evaluating benefit-risk during and beyond drug development: An industry view. *Regulatory Reporter.* 2012, 9 (6).

日本臨牀 72 卷 増刊号 2 (2014 年 2 月 20 日発行) 別刷

最新がん薬物療法学

—がん薬物療法の最新知見—

VII 抗悪性腫瘍薬の臨床試験と開発状況
分子標的薬の臨床試験の現状と課題

OS か PFS か (臨床試験の primary endpoint)

中村 健一 水澤 純基 柴田 大朗 福田 治彦

VII 抗悪性腫瘍薬の臨床試験と開発状況

分子標的薬の臨床試験の現状と課題

OS か PFS か (臨床試験の primary endpoint)

Primary endpoint for cancer clinical trials to establish standard treatment:
Overall survival or progression-free survival

中村 健一 水澤 純基 柴田 大朗 福田 治彦

Key words : overall survival, progression-free survival, true endpoint, surrogate endpoint, hard endpoint

はじめに

従来, がんの第III相試験の primary endpoint の gold standard は全生存期間 (overall survival: OS)であったが, 最近, ① OSは後治療の影響を受けやすい, ② OSの延長に従ってサンプルサイズが増大する, といった問題から無増悪生存期間 (progression-free survival: PFS)が用いられる機会が増えている¹⁾。臨床試験において PFSを primary endpointとして設定しうるか, という問題に対する答えは, 試験の相や目的によって異なりうるが, 本稿では標準治療の確立を目的とした第III相試験において, PFSを primary endpointとしうるかどうかを考えるうえで, 知っておくべき基本的な考え方のフレームワークを述べる。

1) Trueness と hardness

米国 NCIの統計家である Richard Simonは, エンドポイントとは‘患者のベネフィットを測るものさし’であると定義している。この定義に従えばエンドポイントは, 患者のベネフィットを反映し, かつ, ものさしとして正確である必要がある。本稿では前者の‘患者のベネフィット

を反映しているかどうか?’を trueness と呼び, 後者の‘きちんと測れているか?’を hardness と呼ぶ。

1) True 軸

truenessの観点からみた際のエンドポイントには2種類存在する。患者の真のベネフィットを反映する true endpointと, true endpointの‘代替(surrogate)’として用いられる surrogate endpointである。前者の代表は OSで, 後者の代表は奏効割合である。腫瘍縮小自体は患者のベネフィットとはいえないが, 腫瘍縮小が延命につながり, かつ, OSに比べて結果を早くに知ることができることから, 奏効割合は古くから surrogate endpointとして広く用いられてきた。つまり良い surrogate endpointであるためには, ① true endpointとよく相関し, ② surrogate endpointを用いることで何らかのメリットが存在することが必須である。

2) Hard-soft 軸

一方, エンドポイントは‘ものさし’であるため, ものさしとしての機能が優れている必要がある。優れたものさしでは, 誰が見ても同じ結果を導き出せる必要があり, このようなものさしを hardなエンドポイントという。例えば死亡日は誰が見ても変わらないため OSは hardであ

VII

Nenichi Nakamura, Junki Mizusawa, Taro Shibata, Haruhiko Fukuda: Multi-institutional Clinical Trial Support Center, National Cancer Center; JCOG Operations Office/JCOG Data Center 国立がん研究センター多施設臨床試験支援センター, JCOG 運営事務局/JCOG データセンター

るが、QOLなどは人によって評価が異なりうる soft なエンドポイントの代表である。試験の primary endpoint として設定するには (true か surrogate かにかかわらず) hard であることは必須の条件である。

2 Trueness からみた PFS

エンドポイントを考える際に必要なフレームワークとして、trueness と hardness という 2 つの観点を述べたが、この trueness から考えた際には、PFS が primary endpoint として用いることのできる条件は 2 つしかない。 すなわち、① PFS が OS の surrogate endpoint となることが示されている、② PFS 自体が患者にとっての clinical benefit をもつ、のいずれかである。

1) PFS が OS の surrogate endpoint である

メタアナリシスなどで PFS が OS の surrogate endpoint であることが示されている場合には、PFS 自体が clinical benefit をもつことを示す必要はない。これは、surrogate endpoint である以上、PFS が positive であれば、clinical benefit をもつ OS も positive となることが前提であるからである。各種のがんのメタアナリシスで OS に対する PFS の surrogacy を検証しようとする試みがなされているが、現在のところ一定の surrogacy が示されているのは大腸がん²⁾、頭頸部がん³⁾のみであり、反対に胃がん、乳がんなどでは surrogacy が示されていない。

ただ、最近では大腸がんの分子標的薬を含む試験も含めた大規模なメタアナリシスで、PFS が OS の surrogate endpoint とはならなかったという発表がなされており、二次治療以降で有効な薬剤が存在するようながん種では、PFS の OS に対する surrogacy を示すことは今後は難しくなるであろう⁴⁾。

2) PFS 自体が clinical benefit をもつ

PFS を primary endpoint に設定しうる、もう一つの条件は、PFS 自体が true endpoint であること、つまり何らかの clinical benefit をもつことである。PFS ではないが、頭頸部がんにおける喉頭温存生存期間は OS 以外の true endpoint

の好例である。喉頭摘出することになれば、発声、味覚といった点で患者の QOL に大きな障害が生じることから、仮に OS に差が全くないとしても喉頭温存生存期間の延長は患者のベネフィットを直接反映すると考えられる。

一方、通常の PFS がもつ clinical benefit については controversial で、OS が延長するという保証がない状況で、増悪までの期間がわずかに延長したとしても、それは毒性や QOL の低下に見合わない可能性もあり、また同じ 20% の径和増大といっても、小さな病変が 20% 増大したとき、大きな病変が 20% 増大したときでは意味が異なる点も clinical benefit を正確に反映しているとはいえない⁵⁾。clinical benefit をもつかどうか疑わしい状況で PFS が多少延長したとしても、PFS に比べて増悪以降の survival が相対的に長くなってきている現状では、PFS のわずかな延長がもつ意味は更に薄れているといえよう。

以上より、標準治療を決定する第 III 相試験において PFS を primary endpoint として用いるための必須条件は、① PFS が OS の surrogate endpoint となることが示されている、② PFS 自体が患者にとっての clinical benefit をもつ、のいずれかを満たし、かつ、hard であること、となる。NCI の統計家である Korn さんも、'clinical benefit を直接的に測る中間的なエンドポイントが存在しないがん種では、OS を primary endpoint とすべき' と結論している⁶⁾。

3 増え続けるサンプルサイズへの対応策

しかし、現実問題として、OS が 20 カ月以上になるようながん種では、OS を primary endpoint としてランダム化比較試験を組むことは、多くの場合で登録数の観点から困難である。稀少がんなどで十分な患者数が存在しないような場合にも同様の問題が生じるが、このような状況への対応策は、次に挙げるような方策が考えられる。

1) 大規模なサンプルサイズに対応できる体制を構築する

近年、国際共同試験や国内共同試験の必要性が叫ばれているが、共同試験の重要な動機の一つは、OSをprimary endpointとしたときに増え続けるサンプルサイズである。ただし、国際共同試験ではしばしば国別に治療効果の差がみられるというデメリットも存在するため、共同試験の相手方は対象や治療により慎重に検討する必要がある。

2) デザインの工夫によるサンプルサイズの調整

a. 試験の精度を落とす

通常の優越性デザインを取るかぎり、サンプルサイズを減らすために緩められるパラメータは、 α (有意水準)、 β (本当は有効な治療を捨ててしまう確率、 $1-\beta$ が検出力)、 Δ (臨床的に意味のある差)の3つしかない。 Δ は試験治療の毒性などのデメリットに見合ったアウトカムの幅であり、臨床的な意思決定と試験結果を一致させる観点からは Δ を緩めることはお勧めできない。通常の第III相試験では、 α は片側2.5%か両側5%、 β は10-20%に設定されるのが世界標準であるが、対象疾患の患者数が限られている場合には β を20-30%程度、 α についても片側5%や、かなりの稀少がんである場合には片側10%も許容されるかもしれない。もちろん試験の精度を落とすことは望ましくないが、精度を多少落としたとしても試験を行う価値があると考えられる場合には、やむを得ない選択肢とはなりうる。

b. 特殊な非劣性試験

ある特殊な状況では、NCIのFreidlinらが‘hybrid design’と呼んでいる非劣性試験の設定を行うことによりサンプルサイズを減らすことができる場合がある⁷⁾。通常、非劣性試験では標準治療と試験治療で見込まれる予後は同じに設定するが、このデザインでは、試験治療群が少し勝つという見込みがある場合にそれを前提として計画することでサンプルサイズを劇的に減らすことができる。詳細は他稿に譲るが⁸⁾、試験治療が標準治療に比べ有効性以外の何らか

のメリットを有し(非劣性デザインが正当化できる根拠があり)、かつ、試験治療が少し勝つという見込みが根拠に基づき説明可能な場合には適用できる可能性がある。

3) PFSとOSのどちらをもdecision ruleに組み込む

3つめのオプションとしては、PFSでサンプルサイズ設定を行うことによって登録数は現実的な範囲内におさめ、試験のdecision ruleとしては、PFSが有意でかつOSで試験治療が標準治療に一定以上上回る(例:OSのハザード比の点推定値が0.85を下回る=検定は行わない)場合に‘positive’と結論づけるというデザインである。こうすることで、試験治療の真のOSが標準治療の真のOSを下回っている際に、間違っ

おわりに

これまで主にtruenessの観点からPFSがprimary endpointとして用いることのできる条件を述べてきたが、PFSがtrueであれ、surrogateであれ、hardであることは必須条件である。PFS評価には様々なバイアスが入りうるため、これらのバイアスを最小化することはPFSをprimary endpointとして使用する場合には必須である。少なくとも、標準治療を確立する第III相試験では本稿で述べたようなtruenessやhardnessを十分考察のうえで、PFSをprimary endpointとすべきかどうかを決定すべきであり、本稿で示したフレームワークやデザインの工夫がそれらの議論の一助となれば幸いである。

VII