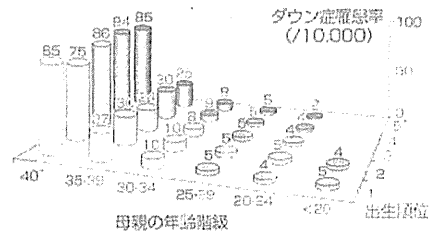
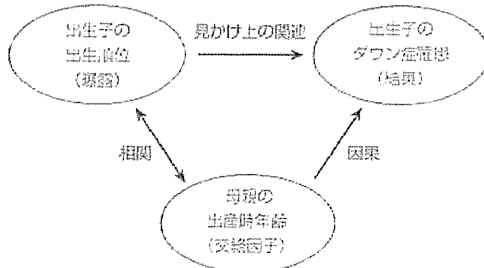


【図3】 出生順位とダウン症罹患率(/10,000)
出生順位とダウン症罹患率の間に関連がみられる。



【図4】 母親の年齢階級別の出生順位とダウン症罹患率(/10,000)
母親の年齢階級別に評価すると出生順位とダウン症罹患率の間に関連はみられない。



【図5】 母親の出産時年齢により生じた交絡の模式図

くなる傾向がある。

特にケース・コントロール研究において、後向きに曝露を評価する場合には、思い出しバイアス(recall bias)の影響に注意する必要がある。後向きケース・コントロール研究における思い出しバイアスとはコントロールに比べてケースの方が過去の曝露情報を正確に記憶しやすい傾向から生じるバイアスである。この傾向から思い出しバイアスはリスクの過大評価につながりやすい。曝露の測定方法の妥当性を事前または当該研究内で検討しておけばその影響の程度も検討することが可能である。

⑤ 交絡と交絡因子

曝露と結果の間の関連を評価する際に、結果に影響を与える他の因子との関連により、曝露が結果に及ぼす影響が見かけ上歪められてしまうことを交絡(confounding)といい、この交絡が生じる原因となる因子のことを交絡因子(confounder もしくは confounding factor)という。交絡の例として、図3、4に出生順位および母親の年齢とダウン症罹患率の間の関連を評価した研究の結果を示した⁴⁾。単に出生順位とダウン症罹患率の関連を評価すると実際に両者に明らかな関連が認められた(図3)ものの、母親の年齢階級ごとに評価すると両者の間の関連はみられなかった(図4)。出生順位とダウン症の間に

観察された関連は母親の年齢を交絡因子とした見かけ上のものであったと解釈できる(図5)。

ある因子が交絡因子となるためには、①曝露と関連すること、②結果に影響すること、③興味のある曝露の結果ではないこと(因果連鎖において曝露と結果の中間にないこと)の3つの条件をすべて満たす必要がある。前述した出生順位とダウン症の例において、母親の年齢は出生順位と相関するとともにダウン症罹患率に影響を与えることから、条件①および条件②を満たすことができる(図4)。また出生順位と母親の年齢には相関があるものの、出生順位が母親の年齢を変えるわけではないと考えられるのが自然であることから条件③も満たされる(図5)。ここで生じた交絡は条件①～③にあてはめることでより明確に整理できる。

⑥ 交絡への対処

データ収集後の統計解析時に事後的に交絡へ対処するよりも、研究計画で事前に考慮する方が一般に好ましい結果につながりやすい。介入試験であればランダム化が唯一最強の解決策であり、非介入試験であれば交絡因子の潜在的な候補をすべて測定できるように計画し、それらを解析に用いることが一般的な解決策となりうる。測定しないことには解析で用いることができず、交絡因子でないことを確認することさえできない。個別の研究の状況に応じて対象者の選択方法や因子の測定計画などを工夫することにより、積極的に交絡に対処することが望まれる。

測定された交絡因子に限り、統計解析時に交絡の調整や交絡に関する感度解析を行うこともできる。交絡を調整するために用いることができる解析法として、限定層別解析、回帰モデルを用いた調整解析、標準化法、傾向スコア(propensity score)を用いた解析などをあげることができる。限定とは結果に影響を与える交絡因子

が同一水準である集団に限定して解析を行うことをいい、層別解析とは限定によって得られる複数の部分集団の結果を統合する解析を一般に表す。前述の出生順位とダウン症の例では、母親のそれぞれの年齢階級に限定した解析により交絡を調整しているものであるといえる。回帰モデルにはロジスティック回帰やCox回帰などが含まれる。回帰モデルを用いた調整解析は、回帰モデルを導入することによって追加されるモデル(仮定)の正しさに依存するものの、より複雑な状況に対して層別解析を一般化したものと考えることができる。標準化法は、曝露の有無の間で比較可能となるように加重平均を用いる方法であり、記述疫学研究分野で古くから用いられてきた方法である。特に近年では、因果推論をもとにする統計手法として傾向スコアを用いた高度な解析なども交絡の調整法のひとつとして頻りに用いられてきており、統計学に関する高度な専門性が疫学研究においても一段と必要とされる時代になりつつある。

文 献

- 1) Last JM: A Dictionary of Epidemiology, 4th ed, Oxford University Press, New York, 2001
- 2) Porta M: A Dictionary of Epidemiology, 5th ed, Oxford University Press, New York, 2008
- 3) Rothman KJ, Greenland S, Lash TJ: Modern epidemiology, 3rd ed, Lippincott Williams & Wilkins, Philadelphia, 2008
- 4) Stark C, Mantel N: Effects of maternal age and birth order on the risk of mongolism and leukaemia. J Nat Cancer Inst 37: 687-698, 1967
- 5) John M. Last(編): 日本疫学会(訳): 疫学辞典 第3版, 日本公衆衛生協会, 東京, 2000
- 6) 坪野吉孝, 久道 茂: 栄養疫学, 南江堂, 東京, 2001

B. 登録・検診・予防

① 癌登録

癌対策を正しく方向付けるには、癌の実態を正確に把握する必要がある。癌登録は癌の実態を把握するための中心的な役割を果たし、癌対策を実施するうえで必須の仕組みである。癌登録には、地域癌登録、院内癌登録、臓器癌登録の3種類がある。罹患率(数)を計測する唯一の仕組みが、地域癌登録である。一方、生存率は、地域癌登録、院内癌登録、臓器別癌登録のそれぞれで計測が可能であるが、目的、対象とする癌患者の範囲、収集する情報がそれぞれ異なる。

② 地域癌登録

地域癌登録は、対象地域の居住者に発生したすべての癌を把握することにより、癌の罹患率と地域レベルの生存率を計測する仕組みである。罹患を把握する主な情報

源は、(1)医療機関からの登録票と、(2)癌死亡情報である。両者を統合して、登録票のない患者の補充登録を行うとともに、登録精度を計測する。同じ患者を誤って複数件計上することのないように、個人識別指標を照らし合わせて個々の患者(腫瘍)ごとに集約する。このための個人識別指標として、正確な生年月日、姓名、性別、住所情報等の収集が不可欠である。一方、対象地域に発生したすべての癌を把握して初めて正確な罹患情報を計測できるので、登録時に個々の癌患者自身から同意を得ることはせずに登録することが国際標準となっている。この点は、個人情報保護の観点からは本人同意原則に對立する行為となり、公衆衛生の観点からの社会全体の利益との調整が必要となる。アメリカ、北欧諸国などでは、法律で癌を届出義務のある疾患と定めることで、この問題を整理している。

わが国の地域癌登録は、2007年度現在35道府県で実施されているが、法的基盤が弱く財政的支援が乏しい中で運営されてきたため、登録精度が低いのが最大の欠点である。世界各国の癌罹患データを収集した「5大陸のがん罹患」最新巻において、わが国から掲載された7登録のみであり、その他の多くの登録は掲載されるための十分な登録精度を達成できていない。作業手順の標準化の促進と登録精度の向上が、わが国の地域癌登録の緊急の課題である。

③ 院内癌登録

院内癌登録は、当該施設で癌の診断・治療を受けた全患者について、癌の診断、治療、予後に関する情報を集約する仕組みである。当該施設における診療の実態を把握し、生存率を計測するなどの機能評価を行うとともに、地域癌登録への届出の役割も果たす。

2002年度から開始された「地域癌診療拠点病院」の指定要件に、「院内癌登録システムの確立または確立の確実な見込み」が盛り込まれ、また、2006年度からは「地域癌診療連携拠点病院」(以下、拠点病院)と名称を変更してその指定要件には、「標準登録様式に基づく院内癌登録を実施すること」が明記された。2008年には、「癌対策情報センターによる研修を受講した専任の院内癌登録の実務を担う者を1人以上配置すること」、および、「毎年院内癌登録の集計結果等を癌対策情報センターに情報提供すること」が明記された。今後、わが国の癌登録を整備するためのいくつかのステップの中で、拠点病院における院内癌登録を整備することが、第一に達成すべきステップであり、これを基盤として、地域癌登録、臓器癌登録の取り組みへと展開することが期待できる。

臨床試験について考える

■ 臨床試験とは

医学の進歩をめざして行う研究のことを「医学研究」と呼びます。医学研究にはいろいろな種類がありますが、それらの中で、治療の開発を目的に人に対して行う研究のことを「臨床試験 (Clinical Trial)」と呼んでいます。ここで「試験」という言葉を用いるのは、臨床試験で行われる医療は日常診療で行われる医療とは異なり、解明されていないことを含むからです。解明されていないことのレベルは臨床試験の内容により異なりますが、いずれの場合も安全に、かつ正しい結果が得られるようしっかりと計画を立てて実施し、段階を追って進める必要があります。臨床試験で扱われる医療行為には、薬以外に放射線治療や手術などがありますが、本章では薬 (抗がん剤) に関する臨床試験を中心に解説します。

臨床試験のうち、薬として販売するための許可を国 (厚生労働省) から得るために行う臨床試験のことを「治験」といいます。治験は薬事法と呼ばれる法律やGCP (Good Clinical Practice)^{*1} と呼ばれる省令などで厳しく規定されています。治験は新薬の製造・販売に関わる企業が行う「企業治験」が一般的ですが、企業が開発に積極的でない領域 (希少な疾患や先端医療研究など) に対しては「医師主導治験」も行われています。治験のほかに薬事法の対象となる臨床研究には、「製造販売後臨床試験」^{*2} や「使用成績調査」などがあります。また、治療成績をより良くするためには新薬の開発だけでなく、承認されている抗がん剤の投与方法の工夫や他の抗がん剤との併用療法、手術や放射線治療との組み合わせなども重要であり、これらの領域に関しては医師自らが研究者となって臨床試験を企画・実施する「医師主導臨床試験」^{*3} が積極的に行われています。

なお、人に対して臨床試験を行うためには、その前段階として安全性や効果について細胞や動物を用いて調べる必要があります。これらを「非臨床試験」と呼びます。非臨床試験はGLP (Good Laboratory Practice) と呼ばれる省令に基づいて実施されます。

【表1】臨床試験の種類

	対象	目的	主体	規範	その他
企業治験	人	未承認薬の承認、適用拡大など	企業	薬事法 GCPなどの省令	法規制対象
医師主導治験	人	未承認薬の承認、適用拡大など	医師	薬事法 GCPなどの省令	法規制対象 企業が積極的でない領域が中心
医師主導臨床試験	人	既承認薬の開発 (併用療法、補助療法) など	医師	臨床研究に関する倫理指針などの指針やガイドライン	法規制非対象



用語のヒント

※1 GCP

正式名称は「医薬品の臨床試験の実施の基準に関する省令」で、臨床試験を受ける人の人権と安全性を確保し、科学的に正しく実施されるために厚生労働省が発した省令です（以前は通達でしたが、1997年より法的拘束力のある省令に改められました）。

※2 製造販売後臨床試験

治験によって安全性と有効性が確かめられて初めてその薬は国から認められ、世に出ることになります。しかし、治験の情報はある一定の条件下で行われたものであり、安全性や有効性に関する情報は十分とはいえません。より多くの情報を得るために、また、より良い治療法を見出すために販売承認を得たあとも臨床試験が行われる場合があり、これを製造販売後臨床試験と呼びます。

※3 医師主導臨床試験

医師が主体となって実施する臨床試験の中で、医師主導治験でないものを本書では医師主導臨床試験と呼びます。医師主導臨床試験は治験と異なり、薬事法やGCPのような省令への遵守は要求されていませんが、「臨床研究に関する倫理指針」などの指針やガイドラインに基づいて実施するよう指導されています。

■ なぜ臨床試験が必要なのか

新しい薬が広く使用されるようになるためには、その薬の安全性や効果を十分に確認することが必要です。有望な薬が見出されたとき、まずは細胞や動物などに対する非臨床試験が実施されますが、得られた結果がそのまま人にあてはまるとは限りません。そのため、非臨床試験のあとには必ず人に対する安全性と効果を確認することを目的とした臨床試験が必要になるのです。

たとえば、手術で取り除くことができないと診断された膵臓がんに対して現在日常診療で使用されているゲムシタピンやエスワンなどの抗がん剤も、以前は膵臓がんに対する効果が明らかでなく、保険適用にはなっていませんでした。しかし、臨床試験や企業治験によって膵臓がんの患者さんに対する安全性と有用性が示されたため、国がそれらの抗がん剤の使用を許可し、日常診療で広く使用されるようになったのです。また、手術可能な膵臓がんに対しては医師主導臨床試験の結果から、術後にゲムシタピンを補助療法^{※4}として行うことが標準治療であるとみなされるようになっていきます。

新しい治療が必ずしも有用な治療であるとは限りません。治療成績を向上させることはそれほど容易ではないため、臨床試験の結果、新しい治療が標準治療にならなかった（新しい治療の効果が証明されなかった）ケースも少なくありません。しかし、臨床試験を実施しなければ、「これまで行われてきた標準治療が引き続き妥当な治療である」という結論も得られないわけですので、このような結果も大切なのです。現在行われている医療は、多くの患者さんにご協力いただいて実施された臨床試験の結果をもとにして成り立っており、将来の医学の進歩のためにも臨床試験は必要不可欠な存在なのです。



用語のヒント

※4 補助療法

手術を受けても再発する可能性のあるがんに対しては、治療成績を向上させるために手術に抗がん剤や放射線治療などを組み合わせることがあり、これを補助療法（アジュバント療法）と呼んでいます。膵臓がんに対する補助療法は手術後に行うことが一般的ですが、最近は手術前に施行する術前補助療法（ネオアジュバント療法）も試みられています。

■ 臨床試験の相（Phase）

新しい有用な治療の開発は、一つの臨床試験を行うだけでは完成しません。有望と考える新しい治療が既に存在している治療に比べ優れているかどうかは、計画的に段階を追って臨床試験を行う必要があり、この治療開発の段階のことを開発の「相（Phase）」と呼びます。実際に治療として使用されるまでには、より実験的な段階から実臨床に近い段階まで順を追って試験を行い、その中で様々な項目を評価することが必要です。

【表2】臨床試験の相

	第Ⅰ相試験	第Ⅱ相試験	第Ⅲ相試験	第Ⅳ相試験
患者数	15～30人	100人未満	一般的に 100人～数千人	数百人～数千人
目的	<ul style="list-style-type: none"> 安全な用量を見つける 薬の投与方法を決める 薬が人体にどのように作用するかを観察する 	<ul style="list-style-type: none"> 薬や治療法が特定のがんに有効性を持つかどうかを判定する 薬や治療法が人体にどのように作用するかを見る 	<ul style="list-style-type: none"> 新しい薬や治療法（新しい使い方）を現在の標準治療と比較する 	<ul style="list-style-type: none"> 新しい薬や治療法の長期の安全性と有効性を評価する

ICR 臨床研究入門のウェブサイト (ICRweb) より改変 <http://www.icrweb.jp/>

第Ⅰ相試験（PhaseⅠ）

第Ⅰ相試験では、新しい薬の安全性の確認と使用法（用量・用法）の決定に重点が置かれます。薬の吸収や分布などを評価する「臨床薬理試験」が行われることもあります。第Ⅰ相試験は通常の薬では健康な人に対して実施されますが、抗がん剤の場合はがんの患者さんが対象になります。

第Ⅱ相試験（PhaseⅡ）

第Ⅰ相試験の結果より、人に投与することの安全性が概ね確認されたと、次はその薬剤が特定のがんに有効かつ安全であるかを調べるために第Ⅱ相試験へと進みます。第Ⅱ相試験は数十名規模の比較的少数の患者さんに投与して行うことが一般的です。

第Ⅲ相試験（PhaseⅢ）

第Ⅱ相試験にて有望であることが示唆された薬は、本当に臨床的に有用であるかどうかを確かめる（検証する）ために、比較的多数の患者さんを対象にした第Ⅲ相試験へと進みます。第Ⅲ相試験では、標準治療※5を対照にして、新しい薬がそれより優れているかどうかを比較検討します。患者さんを標準治療群と新規治療群に無作為（ランダム）

に割り付けることが一般的であり、そのような試験を「ランダム化比較試験」※6と呼びます。抗がん剤の場合、100～300名を対象にした第Ⅲ相試験で承認にいたることもありますが、治療やがんの種類によってはもっと大きな対象者数が必要となります。

第Ⅲ相試験で標準治療よりも優れていることが示され、臨床的意義が認められた薬は、新たな標準治療として日常診療で使用されるようになります（比較的稀ながんに関しては第Ⅱ相試験で良い結果が得られれば使用されるようになります）。臨床試験で試みられる新薬は、安全性と効果のふるいにかけられて淘汰されていくため、第Ⅰ相試験の段階では多数の薬の候補が存在しますが、第Ⅲ相試験を経て新しい標準治療であることが示されるのはその中の一部に過ぎません。

なお、第Ⅲ相試験には、新規治療が標準治療よりも明らかに優れているかどうかを検証する「優越性試験」と、新規治療が標準治療よりも劣っていないことを検証する「非劣性試験」があります。優越性試験は新規治療が標準治療よりも副作用は強いがその分効果も期待できる場合、非劣性試験は新規治療の効果は標準治療と変わらなくても、何らかの利点（副作用が少ない・利便性が良いなど）が期待できる場合に行われることが一般的です。

第Ⅳ相試験 (Phase IV)

市販後に行う第Ⅳ相試験では、新しい治療法の長期的な副作用や有効性の評価を目的とした調査が実施されます。



用語のヒント

※5 標準治療

現時点において最良の治療であることが科学的根拠に基づいて示されており、日常診療で行われることが推奨されている治療のことを標準治療といいます（新しい治療は必ずしも良い治療であるかどうかかわからないので、臨床試験で評価する必要があります。したがって、標準治療が現時点で最も良い治療という位置づけになるのです）。新しい治療はその安全性や有用性を調べる臨床試験で評価され、それまでの標準治療より優れていることが証明されれば、その治療が新たな「標準治療」となります。

※6 ランダム化比較試験

第Ⅲ相試験は新しい治療が標準治療より優れているかどうかを検証することが目的のため、両者を比較することが要求されます。その際に、「バイアス」と呼ばれるデータの偏りを軽減するため、患者さんをランダム（無作為）に割り付けることが一般的であり、このような試験をランダム化比較試験と呼びます。

ランダム化比較試験では、医師や患者さんの主観に基づくバイアスを減らし、より信頼性の高い結果を得ることを目的として、臨床試験で実施する治療の盲検化（ブラインディング）を行うことがあります。医師も患者さんも割り付けられた治療を知られないで実施する試験を「二重盲検（ダブルブラインド）試験」と呼び、患者さんのみが治療を知らされていない試験を「単盲検（シングルブラインド）試験」と呼びます。一方、割り付けられた治療を医師も患者さんも知っている試験は、「非盲検（オープン）試験」と呼びます。盲検化度が高くなるほど、医師や患者さんの主観に基づくバイアス（偏り）が少なくなることから、試験の信頼性が高くなることが期待されます。（表3参照）

【表3】盲検の有無に基づくランダム化比較試験の分類

	臨床試験の間、参加者が 受ける治療について知っている人	
	医師	患者
非盲検試験（オープン試験） 治療に関わる全ての人、どんな治療が行われているかを知っている	○	○
単盲検試験（シングルブラインド） 医師だけが、患者さんがどんな治療を受けているかを知っている	○	×
二重盲検試験（ダブルブラインド） 医師と患者さん両方とも、どんな治療が行われているか知らない	×	×

● 臨床試験の相と病期分類のステージの違い - 膵臓がんの例 -

なお、がん治療の臨床試験に初めて参加された方の中には、「臨床試験の相」と「がんの病期（ステージ）」とを混同される方がいらっしゃいます。ステージは、がんの大きさやリンパ節転移、他の臓器への転移などによって規定される、がんが体内でどれくらい広がっているかを表す指標です。膵臓がんのステージには、国際的に使用されている国際対がん協会（UICC）のTNM分類（図1参照）と、日本で使用されている日本膵臓学会（JPS）（図2参照）による分類がありますので、担当の先生からステージについてお聞きになる場合はどちらの分類に基づいたものなのかを確認されると良いでしょう。

【図1】膵臓がんのステージ分類（UICC分類）

	M0		M1
	N0	N1	Nに関係なく
Tis	0	II B	IV
T1	IA		
T2	IB		
T3	IIA		
T4	III		

分類		分類	
Tis	非浸潤がん	N0	所属リンパ節転移なし
T1	膵内に限局 ≤ 2cm	N1	所属リンパ節転移あり
T2	膵内に限局 > 2cm	M0	遠隔転移なし
T3	膵外に進展	M1	遠隔転移あり
T4	腹腔動脈幹または上腸間膜動脈に浸潤		

UICC：悪性腫瘍の分類 [第7版] より改変

【図2】膵臓がんのステージ分類 (JPS分類) ジャーベ-エス

	M0			M1	
	N0	N1	N2	N3	
Tis	0				
T1	I	II	III	IVb	
T2	II	III	III		
T3	III	III	IVa		
T4	IVa				

分類		分類	
Tis	非浸潤がん	N0	膵臓連リンパ節転移なし
T1	膵内に限局 ≤ 2cm	N1~3	膵臓連リンパ節転移あり (数字はリンパ節群を表す)
T2	膵内に限局 > 2cm	M0	遠隔転移なし
T3	膵内胆管、十二指腸、膵周囲組織に浸潤	M1	遠隔転移あり
T4	隣接する大血管、膵外神経叢、他臓器に浸潤		

日本膵臓学会編：膵癌取り扱い規約 [第6版] より改変

■ 新しい薬が承認され市販されるまでの流れ

以下の図は、新薬が承認されるまでの流れを示したものです。

【図3】新薬が承認されるまでの流れ

非盲検試験	開発段階		適応省令	
創製化合物数 =約30万  承認取得数 = 50	基礎研究 2~3年	新規物質の探索・創薬		
		物理化学的研究		
	非臨床試験 3~5年	薬効薬理試験		G L P
		一般薬理試験		
		一般毒性・特殊毒性研究		
		薬物動態研究		
	臨床試験(治験) 3~7年	第I相試験(臨床薬理試験)		G C P
		第II相試験(探索的試験)		
		第III相試験(検証的試験)		
	承認審査 1~2年	承認申請		
		承認審査		
		医薬品製造(輸入)承認・許可		

ICR臨床研究入門のウェブサイト(ICRweb)より引用 <http://www.icrweb.jp/>

最初に新しい物質（シーズ）の検索や、薬の形にする創薬、物理化学的な性質を調べる基礎研究があり、その後、安全性や効果を細胞や動物に対して検討する非臨床試験が実施されます。これらの過程を経て、ようやく人に対する臨床試験（治験）というステップに進み、数年かけて安全性と効果が調査されます。

治験にて良い結果が出た薬は、いよいよ医薬品としての承認取得へと進みます。日本では、承認のための審査実務は「独立行政法人医薬品医療機器総合機構」※7（以下「PMDA」）という組織が行っています。承認のプロセスは、まず治験の主体者が承認申請者としてPMDAに厚生労働大臣宛の申請書を提出します。提出された書類をPMDAで審査し、その上で厚生労働省の中にある薬事・食品衛生審議会の諮問・答申を経て、厚生労働大臣より薬事法に基づく承認が得られるということになります。薬事法による承認後は、原則として60日以内、遅くとも90日以内に薬価（薬の公定価格）が決まり、医療保険の中で使用できるようになります。

以上のように、一つの薬を育てて診療の中で生かしていくというプロセスは、膨大な費用と手間、時間がかかることがわかりいただけだと思います。



用語のヒント

※7 独立行政法人医薬品医療機器総合機構（PMDA）

厚生労働省所管の独立行政法人であり、英語名称（Pharmaceuticals and Medical Devices Agency）の頭文字をとりPMDAとも呼ばれます。医薬品の副作用などによる健康被害救済業務、薬事法に基づく医薬品・医療機器などの審査関連業務、医薬品や医療機器などの品質を確保する安全対策業務を行っており、日本版FDA※8ともいわれています。しかしながら、FDAに比べてPMDAの職員数は圧倒的に少なく、小規模です。

※8 FDA（Food and Drug Administration）

米国の食品医薬品局のことです。米国食品医薬品局は、日本の厚生労働省にあたる米国の保健社会福祉省（DHHS: Department of Health and Human Services）に属している政府の一機関です。米国内の「食品・医薬品・健康食品・化粧品」等についての厳しい品質検査、および承認審査を行い、消費者保護の権限を与えられています。

● 臨床試験の情報に関して

2005年度より、臨床試験は「臨床試験データベース」に登録されることが推奨されています。日本国内で行われている臨床試験のデータベースには、大学病院医療情報ネットワーク研究センター（通称：UMINセンター）、財団法人日本医薬情報センター（JAPIC）、社団法人日本医師会 治験促進センターなどがあります。これらの情報に関しては、担当医の先生とご相談されたり、セカンドオピニオンを受けたりされると良いでしょう。また、インターネットサービス（独立行政法人国立がん研究センターがん対策情報センターがん情報サービスのウェブサイトの「がんの臨床試験一覧」（http://ganjoho.jp/professional/med_info/clinical_trial/index.html）などを利用されるのも良いと思います。参加できる臨床試験があるかどうかは、膀胱がんのステージや膀胱がんに関する治療歴、全身状態などにより異なりますので、担当の先生によくご確認ください。

（上記で紹介しているサイトは、本書P86～P87をご参照ください。）

■ 代替療法などに関して

がんの治療では通常、手術、放射線治療、化学療法（抗がん剤）が中心的な役割を果たしており、膵臓がんも例外ではありません。しかし、それらの治療には限界があることも事実であり、特に膵臓がんのように有用性が示されている抗がん剤が少ない疾患に関しては、多くの患者さんやご家族が上記以外の治療方法、例えば免疫療法や温熱療法、漢方、健康食品、サプリメントなどを考えられるのではないかと思います。

これらの療法の多くは、これまで述べてきたような臨床試験が過去に実施・報告されていないことから、科学的根拠に基づいて推奨することは難しいレベルにあります。しかし、最近ではこれらの療法においても有効性や安全性を科学的な方法で評価しようという気運が世界的に高まっており、一部の免疫療法などでは膵臓がんに関しても臨床試験が実施され始めています。全ての代替療法に臨床試験という手法がなじむわけではないと思いますが、臨床試験には「将来同じような病気になる患者さんへの財産（次世代への贈り物）」という側面があります。新しい治療が有用であることを証明するためには、大変な労力と時間、費用がかかることがこの章をお読みいただいた方にはおわかりいただけると思いますが、有望と思われる治療に関しては将来像を見据えた開発を行うことが研究にたずさわる者の責務と考えます。

代替療法に関しては、独立行政法人国立がん研究センターがん対策情報センターがん情報サービスのウェブサイトの「がんの治療方法」（<http://ganjoho.jp/public/diagnose/treatment/index.html>）や独立行政法人国立健康・栄養研究所のウェブサイトの「素材情報データベース」（<http://hfnet.nih.go.jp/contents/indiv.html>）、独立行政法人病院機構四国がんセンターのウェブサイトの「がんの代替医療の科学的検証に関する研究」（<http://www.shikoku-cc.go.jp/kranke/cam/index.html>）などもご参照ください。（上記のウェブサイトは、次ページで紹介しています）

〈参考文献〉

本章の作成にあたり、アメリカのNPO法人PanCAN（Pancreatic Cancer Action Network）の医療ブックレット「Clinical Trial」のほかに、以下の文献・資料を参考にしました。

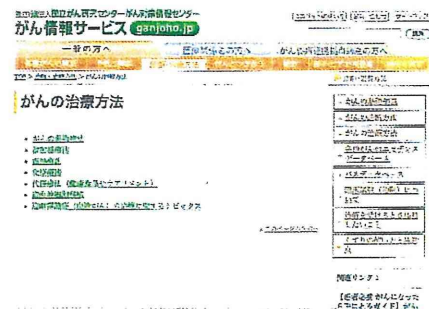
- (1) ICR臨床研究入門のウェブサイト（ICRweb） <http://www.icrweb.jp/>
- (2) 独立行政法人国立がん研究センターがん対策情報センターがん情報サービスのウェブサイト <http://ganjoho.jp/public/index.html>



代替療法について解説されているサイト

国立がん研究センター がん情報サービス「がんの治療方法」

http://ganjoho.jp/public/dia_tre/treatment/index.html



「がん情報サービス」内の「がんの治療方法」のページで、様々ながんの治療法とともに「代替療法（健康食品やサプリメント）」が紹介されています。

国立健康・栄養研究所「素材情報データベース」

<https://hfnet.nih.go.jp/contents/indiv.html>

国立健康・栄養研究所の「健康食品」の安全性・有効性情報」のページの右下に「素材情報データベース」のコーナーが設けられており、素材の名前で検索ができます。



品名	英名	学名	商品番号
アロエ	Aloe (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020101
アロエベータ	Aloe Vera (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020102
アロエベータ抽出物	Aloe Vera Extract (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020103
アロエベータ抽出液	Aloe Vera Juice (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020104
アロエベータ抽出物粉末	Aloe Vera Extract Powder (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020105
アロエベータ抽出物錠	Aloe Vera Extract Tablet (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020106
アロエベータ抽出物カプセル	Aloe Vera Extract Capsule (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020107
アロエベータ抽出物シロップ	Aloe Vera Extract Syrup (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020108
アロエベータ抽出物ゼリー	Aloe Vera Extract Jelly (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020109
アロエベータ抽出物錠剤	Aloe Vera Extract Tablet (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020110
アロエベータ抽出物錠剤	Aloe Vera Extract Tablet (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020111
アロエベータ抽出物錠剤	Aloe Vera Extract Tablet (Lat.)	<i>Aloe barbadensis</i> Mill. or <i>Aloe vera</i> (L.) Bur. & Willd.	20020112

四国がんセンター「がん代替医療の科学的検証に関する研究」

<http://www.shikoku-cc.go.jp/krank/cam/index.html>



左のサイトでは「がんの補完代替医療ガイドブック【第3版】(2012年)」が公開されており、自由にダウンロードができます。



特集

治療効果の判定基準と臨床試験のendpoint

QOL

2) がん臨床試験における QOL評価の問題点*

福田 治彦**

Key Words : quality of life (QOL), health-related quality of life (HRQOL), patient-reported outcome (PRO), clinical trials, oncology

QOL(キューオーエル)

測れるはずがないのに測れると一部の人が信じ、現実的な研究者を辟易させるもので、富山県などで出現すると蜚語とも呼ばれる。

[里見清一 著「誰も教えてくれなかった癌臨床試験の正しい解釈」中外医学社]

はじめに

がん診療・がん研究のコミュニティには、いわゆる“QOL(quality of life)調査肯定派”と“QOL調査否定派”がいる。どちらでもない人の中には“消極的肯定派(雷同派)”と“消極的否定派(懐疑派)”もいるだろう。筆者は里見氏と同じく“積極的否定派”である。

昨今、「がんの臨床試験においては、もはやQOL評価はde facto standard(やって当たり前)である」といったような声を耳にする。本当にそうだろうか？

がん専門医のバイブルといえるDeVitaの教科書¹⁾をみると「HRQOL(health-related quality of life)」の記載は全2,500ページのうち半ページ、

PRO(patient-reported outcome)として約3ページである。がんに詳しい生物統計家として有名なPiantadosiの教科書「CLINICAL TRIALS」²⁾では約600ページ中、PROとして1ページで「quality of life」という項立てすらない。現在筆者らが翻訳中のSWOGのClinical Trials in Oncology第3版³⁾ではQOLは220ページのうち半ページ(18行)であり、SWOGではQOL調査をルーチンには組み込まない理由が書かれている。これら3冊あわせてもたかだか5ページなので読者にも是非読んでみて欲しいが、DeVitaに「今後PROの重要性は増すと期待される」という締めの一文はあるものの、「がんの臨床試験においてQOL評価は当たり前」といった記載は見当たらない。

このように、本稿ではQOL調査に否定的な見解を述べるが、その前提としてまず断っておきたいことは、「(診療や研究において)患者さんのQOLを大事にすること」と「患者自記式の“QOL調査”を行うこと」とは別物であるということである。「QOL調査を行うこと」は必ずしも「患者さんのQOLを大事にする」ことになるとは限らない。後述するようにQOL調査がかえって患者さんのQOLを損うこともあるし、QOL調査否定派の医者が患者さんのQOLを大事にしない医者であるわけでもない(筆者の知るQOL調査否定派/懐疑派の医師は皆、日々の診療において患者さんのQOLを大事にする優れた臨床医である)。

* Dark side of quality of life assessment in oncology clinical trials.

** Haruhiko FUKUDA, M.D.: 独立行政法人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[〒104-0045 東京都中央区築地5-1-1]; Japan Clinical Oncology Group Data Center, Multi-institutional Clinical Trial Support Center, National Cancer Center, Tokyo 104-0045, JAPAN

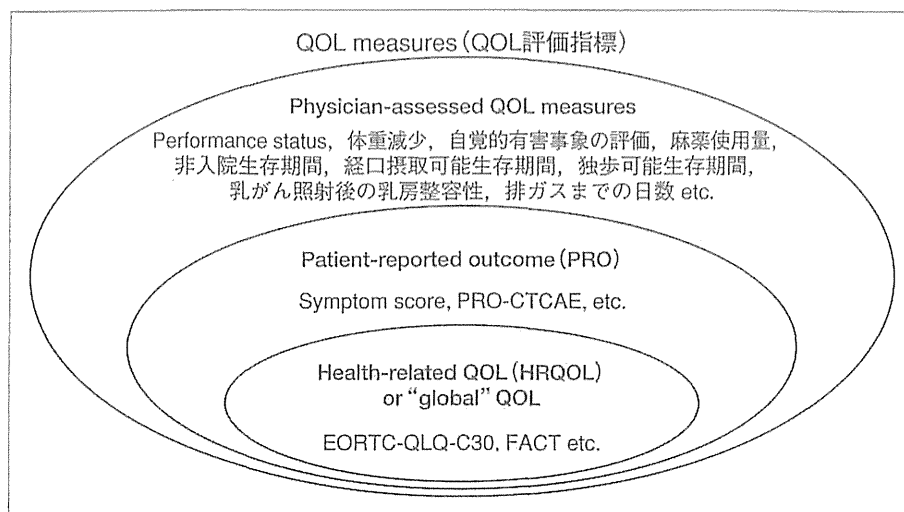


図1 QOL, PRO, HRQOLの関係

「QOL」が論じられるとき、常にこの「QOL」と「QOL調査」の(意図的もしくは無意識の)「混同」が背景にあり、「QOL調査」に対して否定的な意見を述べる者は「患者さんのQOLを大事にしない人間」かのように扱われるリスクを負う。そのためQOL否定派の意見は表に出にくく、QOL肯定派の言のみがさも大勢の世論かのように扱われるという構造的な「非対称性」が存在する。逆に言うと、(意図的か無意識かに依らず)「QOL調査が大事だ」と主張することで、周囲や社会に自分(たち)が「患者さんのQOLを大事にする(善い)人間である」と思わせることができるのである。それは一種の「欺瞞」ではないのだろうか？本稿が、世の中に瀰漫するこうした誤解を解く一助になることを期待する。

QOL (quality of life), HRQOL (health-related quality of life), PRO (patient reported outcome) の関係

「HRQOL」は「PRO」の一種とする考えが一般的である¹⁾⁴⁾。「QOL」を広くとらえた場合、ECOGやKarnofskyのperformance status (PS)等、医療従事者 (health-care staff) が評価するスコアもQOL評価指標 (QOL measures) に含まれる⁴⁾ことから、これらの関係は図1のように整理できる。つまり、「QOL measures」は「physician (またはhealth-care staff)-assessed QOL measures」

と「PRO」に分けられ、さらに「PRO」の一種として「HRQOL」があるという構造である。

筆者が責任者を務めるJCOG (Japan Clinical Oncology Group) データセンターでは、参加施設の担当医やCRC (clinical research coordinator) が記載する通常の症例報告用紙 (case report form; CRF) で収集する有効性や安全性のデータと同様に、担当医やCRCが評価して記載するデータとしての「physician-assessed QOL measures」の開発に積極的に取り組んできた。JCOGデータセンターが考案したエンドポイントには「非入院生存期間 (non-hospitalized survival time) (生存期間-入院日数) や「経口摂取可能生存期間 (ingestion-possible survival)」があるが、これらはPROよりハード (客観的) であり、かつデータ収集に患者さんの負担を伴わないことから、「QOLを評価したい」という研究者の要望に応じて筆者らが考案し、使用してきたエンドポイントである。こうしたハードなphysician-assessed QOL measures はJCOGデータセンターとして今後も積極的に扱う。

また、QOL調査票は、日常診療の一環として、個々の患者さんに症状や日常生活の状況を記入してもらい、その時系列での変化を主治医が把握して治療の意思決定や生活指導に役立てるといった使い方がされることがあるが⁵⁾、筆者はそれについても否定的な見解は持っておらず、むしろ

るそうした臨床医は応援したいと思っている。

以下、がんの臨床試験におけるHRQOLやPROに対する批判的吟味を試みよう。もう一つお断りしておくと、本稿での批判的見解はあくまでも筆者個人の見解であって、JCOGデータセンターの総意ではない。JCOGデータセンターにHRQOL賛成派はいないがPRO賛成派はいる。

HRQOLの臨床的な意味

われわれが、がん治療の臨床試験、特に(QOL評価が重要であるといわれる)第III相試験を行うのは、対象とする患者集団に対して、新しい試験治療が従来の標準治療に比べてよりよい治療かどうかを調べ、試験治療が優れていた場合に「従来の標準治療ではなく試験治療を第一選択とする(新たな標準治療とする)」という意思決定を行い、逆に試験治療が優れていなかった場合には「従来の標準治療が引き続き第一選択である」という意思決定を行うためである。つまりわれわれは、臨床試験の結果に基づいて「治療法の優劣」の意思決定を行う。HRQOLがこの意思決定に役立つのかどうかの吟味において、まずその「臨床的な意味」から問うことにしよう。

がんの標準治療を決める第III相試験では、通常は有効性のエンドポイントである生存期間(overall survival; OS)をprimary endpointとし、安全性のエンドポイントである有害事象(治療との因果関係を問わないあらゆる不都合な事象)や有害反応(治療との因果関係ありと判断される有害事象)をsecondary endpointとして、リスク/ベネフィットバランスの考察を行う。OSに対する代替性(surrogacy)が認められている場合(乳がんや大腸がんの術後補助療法)や、標準治療を決めることを目的としない場合(新薬の製造販売承認を目的とする場合など)には、無再発生存期間や無増悪生存期間をprimary endpointとすることもある。試験治療がOSで勝っていた場合、たとえば、「試験治療は標準治療に対して5年生存割合で10%(に相当する分だけ)上回っていた」時には、「治療を受ける患者さんが100人いるとすると、5年後に生存している人は従来の治療での30人に対して、新しい治療では40人と見込まれます」といった、患者さんにも具体的にイメージできる

情報として説明することができる。毒性についても同様である。「(臨床的に最も重要とされた)grade 3以上の末梢神経障害が20% vs. 30%であった」という結果は、「日常生活に支障がある“しびれ”が100人中20人と30人でした」と、患者さんにも理解可能な情報に翻訳可能である。

一方、HRQOLではどうだろうか? 「QOLスコアが標準治療30点、試験治療40点」であった場合、「患者さんが「生活の質」について付けた点数が標準治療の30点に対して試験治療では40点と勝っていました」といった説明しかできず、この説明を具体的なイメージをもって理解できる患者さんはおそらく存在しない。「治療に満足した人が、100人中30人と40人」という説明なら理解可能であろうが、それではウソになる(満足したかどうかを尋ねているわけではないため)。つまり、HRQOL評価の結果は「臨床的な意味」を持つ情報として提示することはできず、それは有効性や安全性のデータとともにリスク/ベネフィットの考察の組上にあげることができないことを意味する。

また、この文脈で考えた場合、優越性試験と非劣性試験の違いも考慮する必要がある。「毒性が軽い」、「外来治療が可能」等のメリットが新治療にある場合、新治療は有効性(OS等)で勝らなくても、劣ってさえいなければ「よりよい治療」になりうると考えられることから、新治療が(通常OSで)標準治療に劣らないことを検証する非劣性試験(non-inferiority trial)が行われる。非劣性試験におけるリスク/ベネフィットバランスの考察は「有効性は劣らない」かつ「その他の要素で勝っている」のか否かの考察となり、毒性に加えて、もしくは毒性の代わりに「QOLスコアが上回る」を用いることは正当とされうるだろう。なぜなら「QOLスコアが何ポイント勝った」という説明は必要ではなく、「(延命効果は同等でしたが)患者さんが「生活の質」について付けた点数は、新治療が標準治療に勝っていました」という説明は多くの人に何とか理解可能と思われるからである。そして、もし、標準治療と試験治療が有効性でも同等で、(たまたま予想に反して)毒性でも同等という結果であった場合には、QOLスコアが優れた方をよりよい治療とすることは

許容されるだろう(次項で述べる比較可能性の問題は残るが)。ただし、それは有効性で非劣性、担当医評価の毒性でも差がないという限られた状況でのみ有用でありうるのであって、いわば“保険”のようなものである。限られた状況でのみ有用であるものに手間とリソース(および後述する「患者さんへの負担」)をかける価値があるかどうかは常に問われなければならないだろう。

一方、優越性試験(superiority trial)ではどうだろうか？ 試験治療がOSで有意に標準治療を上回った場合、試験治療の毒性が許容可能な範囲にあれば、QOLスコアの優劣がどうであったにせよ、「試験治療が新しい第一選択である」という意思決定は変わらない。逆に、試験治療がOSで勝らなかった場合には、QOLスコアがいくらよかったとしても「(毒性の強い)試験治療が新しい第一選択である」という意思決定にはならないだろう。すなわち、少なくともがんの標準治療を決める優越性の第III相試験において、QOLスコアは治療選択の意思決定に寄与しない。意思決定に寄与しないのであるから、その手間やリソースや患者さんの負担に見合わないとする方が自然な考えであろう。

この、「測っているものの臨床的な意味が説明できない」という問題は、「患者さん自身の評価による“中等度以上のしびれ”は〇〇治療の方が有意に軽かった」といった説明が可能であるPROによる毒性評価には当てはまらず、HRQOL特有の問題といえる。

情報バイアスと比較可能性

次は、「ちゃんと測れているか?」、すなわちバイアスと比較可能性についての吟味である。

世俗的な仮想例で恐縮だが、たとえば、自家用車の販売展示場で行う、試乗後に「乗り心地」を問うアンケートを考えよう。一つは「国産の軽自動車Jにご試乗いただきます」、もう一つは「欧州産の高級車Dにご試乗いただきます」としよう。価格はJが100万円、Dが400万円とし、試乗前に顧客に伝えられることとする。結果、「満足」と答えた人は、Dの40%に対してJが60%と上回っていたとしよう。この結果をもって「JはDより乗り心地がよい」ないしは「JはDよ

り優れた車である」と結論してよいであろうか？ 多くの人の答えは「そんなわけないだろ？」であろう。

別の仮想例を考えよう。ある切除不能がんに対して、標準治療である「薬剤X+cisplatin」と、試験治療である「薬剤X+carboplatin」が非劣性試験デザインでランダム化比較されるとしよう。がん治療の臨床試験では、「cisplatinでは吐き気が強く出ること」と「carboplatinは吐き気が軽いこと」は、試験の意義(rationale)そのものにかかわる情報であるため患者さんへの説明文書には当然書かれなければならない。試験の結果、OSに関して「X+carboplatin」の非劣性が示され、嘔気・嘔吐を含む担当医評価による消化器毒性は期待どおり「X+carboplatin」が軽かった。QOLスコア(もしくはPROの消化器毒性grade)も「X+carboplatin」が有意に良好であった。これをもって「X+carboplatinは(QOLが)よりよい治療である」といえるであろうか？

ここで問題とすべきは、事前に与えられた情報による「情報バイアス」とそれによる「比較可能性」の損失である。車の例と同様に、「+cisplatin」と「+carboplatin」の比較には「比較可能性」が担保されているとはいえない。Jに試乗した顧客が「100万円にしては乗り心地がよい」、Dに試乗した顧客が「400万円にしてはたいしたことがない」と感じるのと同様、「吐き気」に関する事前情報は、患者さんのQOLスコアに影響を及ぼしうる。「吐き気が強い」という事前情報を与えられて「+cisplatin」の治療を受けた患者さんは、自分が「むかむか」を感じたときに「これはcisplatinのせいだ」と感じて低い点数をつけるかもしれないし、逆に「吐き気はほとんど出ない」と説明されて「+carboplatin」の治療を受けた患者さんは「むかむか」しても「治療のせいではないかもしれない。気のせいかもしれない」と考えて点数を低くつけないかもしれない。

全体のQOLスコアに大きく寄与しうる「吐き気」に関する「事前情報」によってQOLスコアは左右されうる。これは疫学でいう「情報バイアス(information bias)」の一種である。患者選択にまつわるバイアスである「選択バイアス(selection bias)」はランダム化により小さくすることがで

表1 検査や診断法の妥当性の評価

		真とする結果 gold standard		尺度
		あり	なし	
検査 結果	陽性	真陽性 true positive	偽陽性 false positive	陽性的中度 positive predictive value (真陽性÷検査陽性)
	陰性	偽陰性 false negative	真陰性 true negative	陰性的中度 negative predictive value (真陰性÷検査陰性)
尺度		感度 sensitivity (真陽性÷あり)	特異度 specificity (真陰性÷なし)	

きるが、情報バイアスはランダム化によっては小さくできないため、サンプルサイズが十分大きいランダム化比較試験であるからといって情報バイアスから逃れられるわけではない(少なくとも臨床医が、ランダム化しているからバイアスは問題にならないと誤解している)。

また、cisplatinと比べてcarboplatinの嘔気が軽いことは確立された知見であるが、開発段階で第II相試験までの毒性に関する情報が必ずしも第III相試験以降で再現されるとは限らない(第II相試験までの毒性データは真の毒性を過大評価もしくは過小評価したものであるかもしれない)ことから、新薬治験の第III相試験においても、情報バイアスによって誤った意思決定につながる情報が生み出される可能性は常にあるといえる。そして、毒性が必発であり、その情報を患者さんに伝え、患者さんの毒性の情報が担当医に適切にフィードバックされることを通じてリスクを最小化することが一般的となっている「がんの臨床試験」においては、こうした情報バイアスは不可避である。この情報バイアスと比較可能性の問題はHRQOLだけではなく、PROによる毒性評価にもあてはまる。

一方、がん以外の臨床試験や、がん患者を対象とはしていても(治療をマスクングすることによる患者さんのリスクの増大がないとみなせる)制吐剤や鎮痛剤の臨床試験では、こうした情報バイアスはプラセボを用いた二重盲検により回避できるため、PROによる毒性評価やHRQOLによる評価は比較可能性が保たれうる。プラセボ対照を用いたランダム化試験は、車の例でいえ

ば、内装と外装を加工して、試乗する車がJとDのどちらであるかわからないようにした上で「乗り心地」を答えてもらい、「どちらの車の乗り心地がよいか」をもって車の優劣をつけることに対応する。QOL調査否定派の言はとかく誤解されがちなので念のために述べるが、筆者は「すべてのQOL評価(やPRO)」に否定的なのではない。「盲検(マスクング)ができないがん治療の臨床試験」において、治療法の優劣を決定する意思決定に用いる評価指標としてHRQOLやPROを用いることは、誤った意思決定を導きうる点において有害ですらありうることを認識する必要があると考えるのである。

QOL尺度の妥当性評価： いわゆる“validation”

臨床検査や診断法の妥当性は、表1に示すように、通常、なんらかの「真」とする結果(たとえば病理診断でのがんか非がん。これを「gold standard」と呼ぶ)を定め、検査や診断に基づく判定(たとえば内視鏡肉眼診断での良悪性)と真の結果との一致/不一致を評価し、真陽性、偽陽性、真陰性、偽陰性の4つの区分から、感度、特異度、陽性的中度、陰性的中度を求めることで評価がなされる。こうした評価を行うことを「妥当性の検討(または検証(validation))」と呼び、従来の検査法や診断法に比べて、新しい検査法や診断法の感度や特異度が十分高い、もしくは遜色がなければ、新しい検査法や診断法の「妥当性が検証(validate)」されたとする。

一方、QOL調査票(自記式アンケート用紙)に

ついて行われる「validation」はこれとは異なる方法による。それは「計量心理学(psychometrics)」的手法によるもので、もともと測れない、すなわち「gold standard」を置くことができない、「心」や「知能」等の評価に用いられる方法論である。計量心理学的な“validation”における指標には、大きく分けて「信頼性(reliability)」と「妥当性(validity)」がある。

詳細は別稿に譲るとして簡単に説明すると、「信頼性」は、「得られる値が安定しているかどうか」の指標であり、「再現性(reproducibility)」と「内的整合性(internal consistency)」からなる。平たくいうと、「再現性」は、「同じ」であることが一貫しているかどうかの指標、「内的整合性」は「違う」ことが一貫しているかどうかの指標、である。一方、「妥当性」は、「測定したいことを本当に測定しているかどうか」の指標であり、「内容的妥当性(content validity：内容が適切か？たとえば痛みを見たいときに痛みを見ているか？)」、「構成概念妥当性(construct validity：他の項目との関係が事前に決めた仮説に沿っているか？)」、「基準関連妥当性(criterion validity：既存の尺度との一致/相関：一致し過ぎてもダメ)」、「応答性/感度(responsiveness/sensitivity：状態の変化に応じて値が変化するか)」からなる¹⁴⁵⁾。

QOL調査票は、こうした信頼性と妥当性が検討されて計量心理学的に“validate”されたものを用いなければならないとされる。こうした“validation”がされていない調査票を用いるべきでないという見解には筆者も賛同するが、逆に、こうして“validate”されたからといって、「治療法の優劣の意思決定に用いること」が妥当であることまで証明されたわけではないことに注意が必要である。つまり、調査票が“validate”されることは「治療法の優劣の意思決定に用いること」についての「必要条件」であるが「十分条件」ではない。「治療法の優劣の意思決定に用いること」が妥当であることの検証がなされたQOL調査票などというものは世の中に存在しない。

ただし、そもそもOSにしろ、担当医評価による毒性にしろ、「治療法の優劣の意思決定に用いること」が妥当であることの検証がなされたエンドポイント自体、世の中には存在しないため、

その点ではQOL調査のみが特別劣っているというわけではない。しかし、「治療法の優劣の意思決定に用いるベネフィット/リスク評価」にOSや毒性を使うことは世界共通の“慣習”であり、それで十分か不十分かの議論はあったとしても、これらを用いることそのものについての異論はないだろう。問題とすべきは、QOL評価を治療法の優劣の意思決定に用いることが万人のコンセンサスになっているわけではないにもかかわらず、ことQOL評価においては「調査票が“validate”されていること」が、あたかも「治療法の優劣の意思決定に用いる妥当性」までが検証されているかのごとき、過大な解釈をされる傾向があることである。「この調査票は“validate”されている」と聞くと、それがなにを意味するのかを踏まえる必要がある。

患者さんの負担(respondent burden)

里見氏も著書で引用している、2008年にJCOに掲載された論文を紹介する⁶⁾。これは、北アイルランドの3つの病院の共同で行われたランダム化比較試験であり、PS 0~2の切除不能肺がん患者を対象とし、対照群の患者には通常の日常診療が行われ、介入群の患者にはEORTC QLQ-C30の日記形式の調査票が渡され、質問に対する回答を16週間、毎週記録することとされた。両群の“QOL評価”にはFACT-LとFACT-Gが用いられた。結果は、primary endpointであったTrial Outcome Indexでは有意差はなかったが(介入群で低く)、secondary endpointsのいくつかの指標では介入群で有意にQOLスコアが低かったというものであった。「Discussion」では、「繰り返し健康状態について回答を迫られたり、日記の調査票を自宅に持ち帰ることで、繰り返し自身の病気についてくよくよ考えたり、心配が増したりしたのではないかと」と考察されている。もっともな見方である。QOL調査が患者さんに苦痛を与える実例といえる。

もう15年も前のことになるが、かつてJCOGの試験で使われていたQOL調査票の質問項目の中には、患者さんが病気になったことを責めるかのような表現や、明らかにデリカシーに欠けると思われる表現がみられた。裏面に「こんな質問

に答えさせられてつらかった」, 「こんな質問になんの意味があるのかわからない」といったコメントが書かれた調査票もあった(当時筆者がJCOGにおけるQOL調査をいったん廃止した背景にはこうした事情もあった)。患者さんと文書で直接コミュニケーションをとるQOL調査やPRO評価にはこうしたリスクもあることを知る必要がある。一方, physician-assessed QOL measureでは, 直接患者さんに「善いこと」もしていないかわりに「悪いこと」もしていない。

QOL肯定派の諸氏の中には, HRQOLやPROの調査を行うことが, 当然「患者さんに善いことをしている」かのようにいう人も少なくないが, 実はかえって患者さんに精神的苦痛を与えている可能性もあることはほとんど語られない。「言語的コミュニケーション」は「両刃の剣」である。すべてのQOL調査において「respondent burden」が慎重に吟味され, 最小化されなければならない。

欠測値の問題

QOL評価における「欠測値(missing data)」の問題は広く知られており, DeVitaにも「analytic consideration」として取り上げられている¹⁾。つまり, QOL評価における欠測値はランダムに発生するのではなく, 患者さんの状態が悪くなったがゆえに生じた欠測であることが多く, 欠測値を解析から単純に除外してしまうことでQOLの値を過大評価してしまうことになるという問題である。原病の増悪や死亡, 治療の毒性のために調査票の記載ができない健康状態になってしまうことにより欠測が生じる。完全にランダムに生じた欠測(missing completely at random)なら解析から除外してもバイアスは生じず, そうした欠測は「uninformative missing」と呼ばれるが, QOL評価における欠測の多くはそうではないため「informative missing」と呼ばれ, バイアスを小さくするための解析上の工夫がなされなければならない。具体的には, 欠測値に, 最悪値(ゼロ点)を当てはめたり, 欠測の前の直近の値を当てはめたり(last observation carry forward approach), 回帰モデルによる予測値を当てはめたりする方法(imputation approach)など

があるが, いずれも「真の値」を用いるわけではないため限界があり, どの処理が最適かに関するコンセンサスはない。DeVitaでは「調査期間を長く取り過ぎない」という, 解析以前のデータ取得レベルでの現実的な対応を推奨している。

PROに関するFDAガイダンス

2009年に米国FDA(Food and Drug Administration)が公表したPROに関するガイダンス「Guidance for Industry-Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims」²⁾が, あたかも「FDAがPROを公認した」証拠かのように言われることがあるが, それは事実と反する。

このガイダンスでFDAは, 「QOLやPROの改善を添付文書で謳う医薬品の効能効果(labeling claim)に含めるのであればこのガイダンスに従う必要がある」と述べているにすぎず, 「医薬品の薬事承認の審査に(HRQOLを含む)PROの結果を用いるべき」や, 「PROを推奨する」といったことは一言も書かれていない。むしろ, マスキングしていない試験におけるPROに対する否定的見解が明記されており(open-label clinical trials are rarely adequate to support labeling claims based on PRO instruments), このガイダンスによれば, FDAはむしろ「マスキングできないがんの臨床試験におけるPRO」に対しては否定的と考えるのが正しい理解であろう。このことは, DeVitaにも「the U.S. Food and Drug Administration (FDA) is reluctant to consider claims about HRQOL improvement in approval decisions」と言及されている³⁾。

「日本語」の問題

先述の「respondent burden」に絡めて「日本語」の問題に触れる。HRQOLやPROの質問文の日本語に不自然さを感じたことのある人は少なくないであろう。これには一応理由がある。

QOL調査票のオリジナル版は多くが英語であり, そのため日本を含め, 英語を母国語としない国におけるQOL調査では, 当然のことながら使用国の言語に翻訳したQOL調査票が用いられる。しかしオリジナルの英語版作成者にとって

は、せっかく「validate」された調査票を勝手に不適切に翻訳して使われては困るので、QOL調査の質問内容は使用国の言語に順翻訳(translation)されたのち、それを英語に逆翻訳(back translation)し、逆翻訳した英語とオリジナルの質問文を対比して「同じことを尋ねている」ことを確認する、という作業が通常要求される。QOL調査票の質問の日本語の不自然さは、(少なくとも幾分か)この「translation-back translation」のプロセスに起因する。しかし、「不自然」だけでは問題とするには当たらないだろうが、なかには「不自然」では済まず、日本語に敏感な患者さんだったら「人を(小)馬鹿にしている」と感じるのでは?と思われる質問文をみることが少なくない(少なくとも自称「日本語にうるさい」筆者はしばしばそう感じる)。「人を小馬鹿にした」かのような日本語の質問票は、QOL調査に協力して下さる患者さんの精神的苦痛(mental respondent burden)になりうる(イヤな感じを抱きつつ質問に答えさせられることが苦痛であることに異を唱える人はいまい)。

そもそも、聖書を持つキリスト教文化の英語圏と、聖書に相当する文書を持たない日本文化では「文書」が持つ意味自体が違うし、「書かれたものがすべて」の米国の契約社会文化と、「行間を読む(それを言っちゃあおしめえよ)」、「縁起の悪いことは言わない/書かない(言霊思想)」、「甲乙両者が誠意をもって対応(和をもって尊しとなす)」の日本文化とでは「言語的コミュニケーション」の意味やあり方も異なる。自分の意見を明言することに関する態度も違うではないか(自分の意見を言わないと叱られる米国 vs. 自己主張し過ぎると嫌われる日本、first nameで呼び合う米国 vs. 本名は呼ばず職名で呼ぶ日本)。さらに、主語の扱いがまったく異なるという文法の違い⁹⁾等々の彼我の差を考えれば、オリジナルの英語のQOL調査票の質問を、back translationを経てもなおオリジナルの英語と同義とみなされるように、かつ日本語として自然で適切な(失礼でない)質問に翻訳する、という作業には、卓越した日本語能力と語学センスが必要であろう。

難しいことはわかるし、「back translationを前提とするため不自然な日本語になるのはやむを

えない」という言い訳を全否定するつもりはないが、QOL調査票やPROが「言語的コミュニケーションツール」である以上、少なくとも、調査に協力して下さる患者さんが読んで不愉快に思う「小馬鹿にしたような表現」になっているか否かを見極める程度の日本語能力は必要であろう。必要な語学力のレベルと文化の違いを考えると、この「translation-back translation」という方法自体にそもそも無理があるのではないかと筆者は考える。オリジナルの英語とback translationした英語を比べるのではなく、たとえば、オリジナルの英語と翻訳した日本語を(英語日本語双方に精通した優秀な翻訳家などの)第三者がチェックするといったような“validation”を英語圏のQOL研究者に飲ませられないものだろうか?もともときわめて“ソフト”な計量心理学である。ここだけ厳密にしても仕方なかろう。ちなみに、“I love you”を、夏目漱石は「月がキレイですね」と訳し、二葉亭四迷は「わたし、死んでもいいわ」と訳したらしい。「translation-back translation」の枠組みでは漱石も四迷も確実に“失格”である。

おわりに

エンドポイント(endpoint)は「患者さんのベネフィットを測るものさし(criterion by which patient benefit is measured): by Richard Simon」であり、「ハード」なエンドポイントと「ソフト」なエンドポイントがある。「ハード」とは「誰が見ても同じ・何回見ても同じ」ことを意味し、「ソフト」とは「見る者によって評価が異なる・時々によって違う・他の影響を受けやすい」ことを意味する。HRQOLやPROが「ソフト」なエンドポイントであることは、QOL調査否定派かQOL調査肯定派かに依らず共通認識であろう。そして、研究者個人や研究組織として、「ハード」なエンドポイントを優先する立場と「ソフト」なエンドポイントを優先する立場がありえ、前者は「conservative」なスタンス、後者は「liberal」なスタンスともいえる。

「conservative」なスタンスとは、「本当はよくないものを誤ってよいと判断する偽陽性の誤りを避ける」ことを優先する立場/嗜好/方法であり、確率的な α エラーの最小化を優先すること

に対応する。「liberal」なスタンスとは、「本当はよいものを誤ってよくないと判断する偽陰性の誤りを避ける」ことを優先する立場/嗜好/方法であり、確率的な β エラーの最小化を優先することに対応する。偽陽性の誤りと偽陰性の誤りの両方とも最小化できればよいのだが、個々の試験のデザインや結果の解釈等においては、この「conservative」と「liberal」はトレードオフの関係にあり、残念ながら「非常にconservativeであり、かつ非常にliberalである」ことはできない。

QOL調査肯定派か否定派かの違いは、ソフトなエンドポイントを重視または許容する「liberal」なスタンスか、ハードなエンドポイントを重視する「conservative」なスタンスかの違いであるともいえる。世の中の方法論がすべて「conservative」であれば、本当はよくない治療を誤ってよいとしてしまうリスクは小さくなるが、本当によい新しい治療が日の目をみないことによって医学の進歩は遅くなるかもしれない。逆に、世の中がすべて「liberal」ならば、医学の進歩の効率はよいかもしれないが、本当はよくない(場合によっては害のある)治療を受けるという国民の不幸やリスクは増える。つまり、これはどちらが正しくてどちらが間違っているという問題ではないのである。もちろん「善悪」の問題でもない。

筆者は、「QOL調査は善である」といった主張には反論するが、だからといって「QOL調査は悪だからやるべきでない」というつもりもない。要は肯定・否定は立場/嗜好/方法の違い、平たくいえば「好み」の違いであるということである。そして、好みの問題なのだから、QOL調査はそれを好きな人が、賛同してくれる人を語らって(患者さんにも迷惑をかけないように)やればよいと考える。その際、「やるべきものだから」や「de facto standardだから」や「FDAも認めているから」などの(おそらくは正しくない)根拠を基に、それがキライな人たちにまで押しつけないで欲しいと思っているだけである。

筆者らはQOL“調査”がキライなだけであって、「患者さんのQOLが大事」という信条はQOL肯定派の諸氏と変わりはないと思っている。筆者らは筆者らが正しいと思うやり方で「患者さんのQOLをよくする治療」の治療開発・臨床試験に携わっ

ており、個々の臨床試験において、どうやったら患者さんのリスクをもっと最小化できるのか・・・等を日々考えている。QOL調査よりも優先すべき未解決の課題が「がんの臨床試験」にはまだまだあると筆者らは思っているだけである。

最後の最後においてお願いで稿を終えたい。どうかQOL調査否定派の私たちのことは放っておいていただきたい。そして「QOL調査」に非協力的な私たちのことを悪人呼ばわりしないで欲しい(そういう「空気」も醸成しようとしなくて欲しい)。QOL調査をやるなら、QOL調査が好きな人たちだけでやって欲しい。そうすれば私たちも、他人が好きでやっていることにケチをつけるようなオトナげないことはせずにすむのだから・・・

補遺：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23-A-16(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) Earle CC, Schrag D. Health Services Research and Economics of Cancer Care. In : DeVita VT, Lawrence TS, Rosenberg SA, editors. Cancer- Principles & Practice of Oncology. 9th edition. Philadelphia : Lippincott Williams & Wilkins ; 2011. p. 352.
- 2) Piantadosi. CLINICAL TRIALS- A Methodologic Perspective. Second Edition. Hoboken : John Wiley & Sons, Inc. ; 2005. p. 207.
- 3) Green S, Smith A, Benedetti J, Crowley J. Clinical Trials in Oncology. Third Edition. Boca Raton : CRC Press ; 2012. p. 45.
- 4) Walters SJ. Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation. Chichester : John Wiley & Sons, Ltd. ; 2009. p. 2.
- 5) Walters SJ. Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation. Chichester : John Wiley & Sons, Ltd. ; 2009. pp. 31-49.
- 6) Mills ME, Murray LJ, Johnston BT, et al. Does a patient-held quality-of-life diary benefit patients with inoperable lung cancer?. J Clin Oncol 2008 ; 27 : 70.
- 7) U.S. Department of Health and Human Services- Food and Drug Administration. Guidance for In-