代表的な治療領域は，いわゆる見捨てられた疾患（neglected disease）と希少難治性疾患（orphan disease）である．

見捨てられた疾患とは，主として開発途上国において大きな負担となっているマラリア，アフリカトリパノソーマ病，リーシュマニア病などの熱帯病である．これらの疾患群は，患者が多数であるにもかかわらず，患者に購買力がなく市場性に乏しいことから，医薬品の研究開発が行われないことを特徴とする．これらの疾患は先進国では大きな問題とならないことから，伝統的に国家による研究助成が行われておらず，製薬企業がこれらの疾患治療薬の研究開発に取り組むためのインセンティブが与えられていないことが要因として指摘されている[2]．

希少難治性疾患とは，きわめて患者数が少なく，かつ有効な治療法が存在しない疾患を指す．見捨てられた疾患同様，市場性に乏しいために医薬品の研究開発が滞ることが問題となる．これらの疾患群におけるアンメット・メディカルニーズの充足のためには，研究開発を行う企業や団体へのインセンティブの仕組みの構築が重要となると考えられる．

4. アンメット・メディカルニーズが問題となるその他の疾患

上述の見捨てられた疾患は，主として開発途上国において問題となる疾患群であるが，2004年のWHOの報告は先進国を含めた世界全体において疾病負担の大きい慢性疾患にも焦点を当てている[2]．世界における死因の第1位は虚血性心疾患，第2位は脳卒中であり，今後，開発途上国においてこれらの疾患による死亡数は著しく増加すると予想されている．これら慢性疾患におけるアンメット・メディカルニーズに対応することは，先進国のみならず開発途上国にとっても大きな利益をもたらすと考えられる．

WHOにより優先順位が高いとされたその他の疾患領域には以下のものがある．①薬剤耐性菌の拡大とインフルエンザパンデミック：将来に重大な結果をもたらす可能性のある疾患．②心臓発作と脳卒中の二次予防：負担が大きく予防可能な疾患であり，固定配合剤の開発等により大きな改善が見込まれる疾患．③変形性関節症とアルツハイマー病：負担が大きく治療法がない疾患．④がんと糖尿病：負担が大きく既存治療では不十分である疾患．

5. 世界におけるアンメット・メディカルニーズへの対策

世界においては，見捨てられた疾患の治療薬の研究開発促進を目的とし，Public-Private Partnership（PPP）によるいくつかの活動が行われている．これらのPPPはProduct-Development Partnership（PDPs）と呼ばれ，Drugs for Neglected Diseases（DNDi）の活動がよく知られている[3]．DNDiは，世界各国の研究所，NGO，規制当局が共同設立した国際的ネットワークであり，市場性ではなくニーズ最優先で新薬開発を行う．現在までに抗マラリア合剤2薬剤の上市に成功しており，各国政府への働きかけも積極的に行っている．また，WHOの協力を得た国際的ネットワークであるGlobal Forum for Health Researchは，さまざまなPDPsの支援活動や情報提供活動を行っている[4]．

見捨てられた疾患の治療薬開発に対するインセンティブに関しては，2007年に米国Food and Drug Administration（FDA）がFood and Drug Administration Revitalization Act（FDA再生法）改訂において打ち出したPriority Review Voucher（PRVs）が注目される[5]．これは，見捨てられた疾患の治療薬を上市した企業に対し，他の薬剤の承認申請時にPriority Reviewを受ける権利を与えるものである．Voucherは薬剤の種類を問わず使用でき，第三者への譲渡も可能である．通常のインセンティブが有効でない薬剤の開発を促進するための新たな発想である．

世界の貧困と不公平の解決に向けて活動する

NGOであるOxfam Internationalは，2008年の報告書（Oxfam Briefing Paper）のなかで，上述のさまざまな活動を評価し，見捨てられた疾患治療薬開発への投資はいまだ不十分であり，医薬品開発体制やインセンティブの仕組みに関する改善が必要であることと指摘している．詳細は本報告書を参照されたい[6]．

希少難治疾患治療薬については，各国において研究開発促進策（オーファンドラッグ制度）が整備されている．希少性の定義は国により異なるが（米国20万人未満，EU 5/1万人未満，日本5万人未満），インセンティブとして各国共通にあげられているものは，研究計画立案に際しての規制当局の支援，経済的メリット（税制上の優遇措置や開発費用の還付），審査上の優先措置などである[7~9]．

6. 日本におけるアンメット・メディカルニーズの現状と対策

WHOが指摘するアルツハイマー病，がん，糖尿病，希少難治性疾患などは，日本においてもアンメット・メディカルニーズの高い治療領域としてしばしば言及される．財団法人ヒューマンサイエンス振興財団による2010年の医師アンケートに基づく，治療満足度と薬剤の貢献度との関係を図2に示す[10]．消化性潰瘍，高血圧症，高脂血症などは，治療満足度と薬剤貢献度がともに高い疾患群であるが，アルツハイマー病，糖尿病合併症などは治療満足度，薬剤貢献度ともに低い疾患群である．がん領域では治療満足度に幅があるが，薬剤貢献度は一様に低く評価されている．

なおこの調査は，第1回：1995年，第2回：2000年，第3回：2005年，第4回：2010年に行われており，歴史的な変化を見ることができ，貴重なものである．患者数やQOLと組み合わせ「疾病負担」のコンセプトとリンクできるようになることが望まれる．

日本においては，海外では標準薬として使用されている医薬品が日本では承認されていない「ドラッグラグ」による治療上のギャップも，アンメット・メディカルニーズとの関連で議論されている．筆者らの研究では，1999年から2007年の間に米国，EU，日本のいずれかで承認された新有効

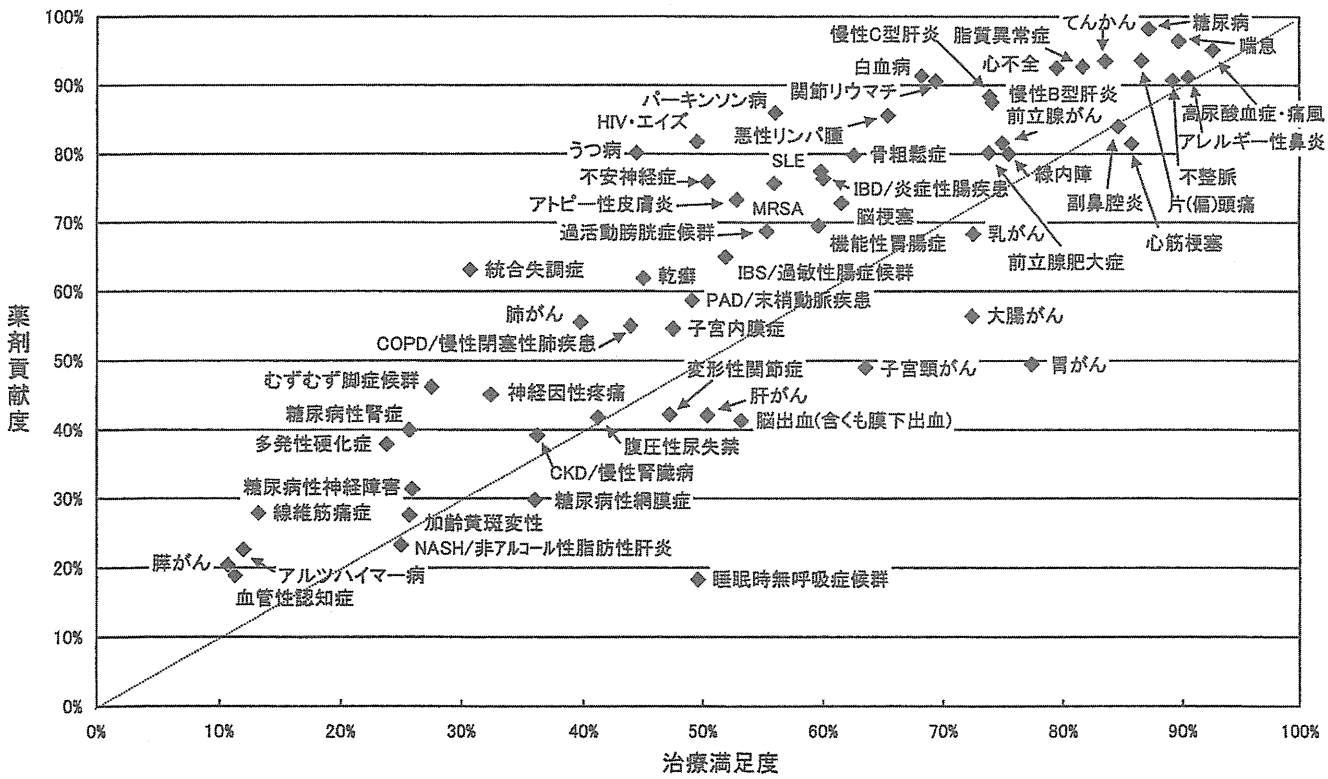

図2　治療満足度と薬剤貢献度との相関図（2010年度，ヒューマンサイエンス振興財団の調査）[10]

成分含有医薬品398薬剤のうち，2007年末時点において日本で承認されていた医薬品は220薬剤（55.3%）にすぎず，これらの日本での承認は海外初承認から約3.5年遅れていた（中央値41.0カ月）．臨床的重要度が高い146薬剤（既存治療に比べて明らかに高い有用性を有するとして日米EUいずれかにおいて審査上の優遇措置を受けたもの）を抽出して分析すると，日本での承認は72薬剤（49.3%）と半数を下回った．重要度が高いにもかかわらず日本で開発が行われていないものは，希少がんや先天性代謝異常症などの希少難治性疾患であり，患者数がきわめて少なく市場が小さい疾患領域で開発着手が滞ることが明らかとなった[11]．

厚生労働省は，ドラッグラグの解消に向け，2005年以降，いくつかの施策を講じている．「未承認薬使用問題検討会議」および本検討会議を引き継ぐ形で設置された「医療上の必要性の高い未承認薬・適応外薬検討会議」では，医療上の必要性が高いと判断された未承認薬について，企業への開発要請が行われている[12,13]．また，これらとの関連で，薬価制度上の施策「未承認薬・適応外薬等解消促進加算」が2010年から試行的に導入された．これらの施策の実効性については今後の検証が必要であろう[14]．　　[辻　香織・津谷喜一郎]

文　献

1) 津谷喜一郎編：くすりギャップ：世界の医薬品問題の解決を目指して．ライフサイエンス出版，2006.

2) 川島今日子，川上純一，津谷喜一郎（訳）：ヨーロッパと世界のプライオリティ医薬品―エグゼクティブサマリーの日本語訳―．薬理と治療，8: 803-812, 2005. In 津谷喜一郎（編）：くすりギャップ：世界の医薬品問題の解決を目指して．ライフサイエンス出版，2006，pp. 65-74.

3) Drugs for Neglected Diseases. Available from URL: http://www.dndi.org/ [Accessed Apr 12, 2010]

4) Global Forum for Health Research. Available from URL: http://www.globalforumhealth.org/Site/000__Home.php [Accessed Apr 12, 2010]

5) FDA. Amendment to the Food and Drug Administration Revitalization Act. 2007. Available from URL: http://www.fda.gov/oc/initiatives/HR3580.pdf [Accessed Apr 17, 2010]

6) Oxfam International. Ending the R&D Crisis in Public Health: Promoting pro-poor medical innovation. Oxfam Briefing Paper 122. 2008 Available from URL: http://www.oxfam.org/sites/www.oxfam.org/files/bp122-randd-crisis-public-health.pdf [Accessed April 12, 2010]

7) FDA. Orphan Drug Act (as amended) Available from URL: http://www.fda.gov/orphan/oda.htm [Accessed Apr 17, 2010]

8) EMEA. Orphan drugs and rare diseases at a glance. London; 2007 Available from URL: http://www.emea.europa.eu/pdfs/human/comp/29007207en.pdf [Accessed Apr 17, 2010]

9) 医薬基盤研究所：希少疾病用医薬品開発振興業務．Available from URL: http://www.nibio.go.jp/shinko/orphan.html [Accessed Apr 17, 2010]

10) ヒューマンサイエンス振興財団：平成22年度国内基盤技術調査報告書―2020年の医療ニーズの展望―．

11) 辻　香織：日本におけるドラッグラグの現状と要因．薬理と治療，37: 457-495, 2009. Available from URL: http://www.lifescience.co.jp/yk/jpt_online/review0906/index_review.html [Accessed Apr 20, 2010]

12) 厚生労働省：未承認薬使用問題検討会議 Available from URL: http://www.mhlw.go.jp/shingi/other.html#iyaku [Accessed April 20, 2010]

13) 厚生労働省：医療上の必要性の高い未承認薬・適応外薬検討会議．Available from URL: http://www.mhlw.go.jp/shingi/other.html#iyaku [Accessed April 20, 2010]

14) 厚生労働省：第158回中医協資料．Available from URL: http://www.mhlw.go.jp/shingi/2009/12/dl/s1222-5b.pdf [Accessed April 20, 2010]

C14 薬剤経済学

薬剤経済学（pharmacoeconomics）は，くすりの費用対効果（cost-effectiveness）を評価する学問をいう．医薬品を開発し，市場に出すのに際しては，1900年代の「品質」（quality）にはじまり，1930年代からは「安全性」（safety），1960年代からは「有効性」（efficacy）のエビデンスが求められてきた．さらに1990年代初頭から，限られた医療資源の合理的使用という観点のもと，医薬品の「費用対効果」すなわち「効率」（efficiency）のエビデンスが，市場における価値を決める際に重用視されるようになってきた．具体的には薬の価格決定や，保険償還の可否の決定の際に，判断基準の一つとして薬剤経済学的データが利用されている．

1. 薬剤経済学の本質

薬剤経済学のゴールは，「医薬品に，価値に見合った効果があるか？」（value for money）を評価することにある．すなわち，新たな医薬品を導入することで既存の医薬品と比較して「どれだけコストが変化するか？」「健康アウトカム（health outcome）がどれだけ変化するか？」の2点について，双方のバランスを評価して，医薬品の価値を明らかにすることが目的である．なお「健康アウトカムの向上」には，「血圧を低下させる」「心筋梗塞の発症を減らす」「生存年数を延長する」「生存者数を増加させる」など多種多様の概念が含まれる．以降では，健康アウトカムを単純に「アウトカム」と表記する．

そのため，コストのみを比較した解析や，既存の医薬品などのコントロールをおいた比較を行っていない解析は，完全な薬剤経済評価（full economic evaluation: FEE）ではなく，部分的な評価（partial evaluation）に分類される[1]．前者はアウトカムの向上について情報がなく，後者は「コントロールと比べてどれだけコスト・アウトカムが変化するか」の情報がないため，正しい評価を下せない．

医薬品の価値を正しく評価するためには，「新しい医薬品の導入に必要なコスト（介入のコスト）」，「介入によって，将来削減しうる医療費などのコスト」，「介入の導入によるアウトカムの改善度合」の3点を定量的に見積もる必要がある．

「薬剤経済学的にすぐれている」ことと，「将来の医療費の削減幅が，介入のコストを上回ること」とは，まったく別物である．実際新規の医薬品についてコスト面だけを評価した際に，総コストが既存の医薬品よりも低くなるものはきわめてまれである．

2. 薬剤経済評価の基本的概念—増分費用効果比の考え方—

単純化した薬剤経済評価の概念をコスト-効果平面図（図1）に示す．評価の基本は，コントロールのコストとアウトカムを中心に，そこからの差を考えることになる．「費用対効果」という言葉からすぐに連想されるのは，図1左隅の「原点」から延びた2本の直線だが，実際に評価すべき

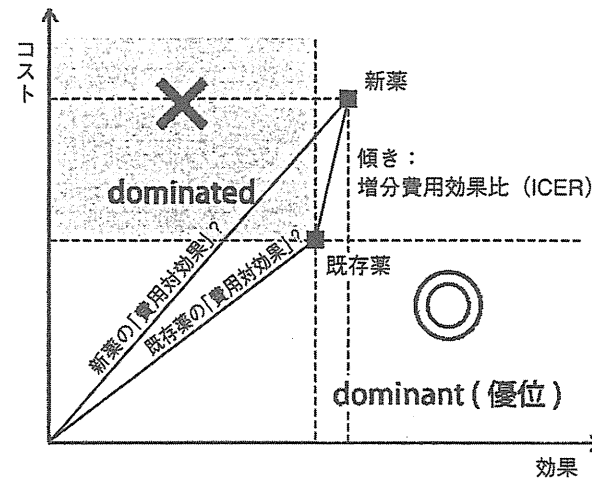

図1 コスト-効果平面のなかの増分費用効果比

は，既存薬と新薬を結ぶ直線である．具体的には「増分費用効果比（incremental cost-effectiveness ratio：ICER）」を計算して評価する．ICERの値は，新薬とコントロールである既存薬のコストの差とアウトカムの差を算出した上で，前者を後者で割ることで求められる．すなわち，図中の太線の傾きがICERになる．

ただし，図1右下の「コストを削減し，アウトカムを改善できる場合」と，左上の「コストは増大する上に，アウトカムは悪化する場合」については，あえてICERを求める必要はない．前者なら「導入すべき」，後者なら「導入すべきでない」という，自然な結論になる．前者のように，コントロールと比較してコスト削減・アウトカム改善になる医薬品のことを「優位（dominant）の状態にある」と表現する．

3. 薬剤経済評価の四つの手法とアウトカムの扱い方

この項では，具体的な経済評価手法を三つに分けて論ずる．三つの手法の違いはアウトカムの測り方で，コストの算定法はすべて共通である．コストについては，次項で詳しく扱う．いずれの手法も，「コストの増加分に見合った，アウトカムの改善があるかどうか？」を評価する点では一定である．

まず，費用効果分析と費用効用分析について述べる．この二つの手法は，いずれも先ほど触れたICERを算出して評価を行うものである．

例として，急性心筋梗塞についての血栓溶解剤の経済評価をとって説明しよう．新薬と既存薬があり，コストについては新薬が1人あたり10万円，既存薬が1人あたり2万円とする．アウトカム指標としては，救命人数をとる．新薬では100人中93人が助かり（7人死亡），既存薬では100人中90人が助かる（10人死亡）ものとする．また簡単のため，両群ともに100人の患者に投薬すると考えよう．コストに関しては，新薬群が10万×100＝1000万円，既存薬群が2万×100＝200万円となる．一方，アウトカムは，新薬群では93人が，既存薬群では90人が助かることになる．

結果を統合すると，新薬は800万円の追加費用で救命人数を3人増やせることになる．すなわち救命人数1人増加あたり267万円になる．これが「アウトカム1単位増加あたりの費用」すなわちICERである．このような分析を費用効果分析（cost-effectiveness analysis：CEA）と呼ぶ．あくまで全体での比較であり，「1人に267万円つぎ込めば必ず救命できる」と解釈してはならない．

アウトカム指標として救命人数をとった場合，生命予後への影響が評価しづらい領域，たとえば慢性疾患の医薬品との比較は難しくなる．時間的な制約などから，より測定が容易な血圧やコレステロール値，心筋梗塞の発症など，疾患特異的なアウトカムを用いる場合もあるが，このような場合は他の疾患領域の医薬品との比較は完全に不可能となる．また，疾患発症後の生活の質の低下は，生存年数を物差しにしても評価できない．関節リウマチや脳梗塞，認知症などの原因で，介助者なしでは外出できない状態で1年生きるのと，完全に健康な状態で1年生きるのとでは価値は変わってくるだろう．

こうした考え方を発展させたのが質調整生存年（qualty-adjusted life years：QALY）の概念である．QALYの算出に際しては，特定の健康状態に「効用値（utility score）」をあてはめる．効用値は，死亡が0，完全に健康な状態が1となる．骨破壊が進行し，介助者なしには外出ができない状態（状態Aとおく）に0.4を当てはめたとしよう．すると状態Aで1年生きることは，生存年数では当然1年だが，QALY基準では1×0.4＝0.4QALYに換算される．状態Aで10年生きることと，完全に健康な状態で4年生きることとが，どちらも4QALYで同等となる．

QALYとLYとQALY gainedの関係を，図2に示した．点A（時間 time：t_1）で疾患にかかり，点B（t_2）まで徐々に効用値が低下していく．そして点Bで，介入を選択することになる．無治療の場

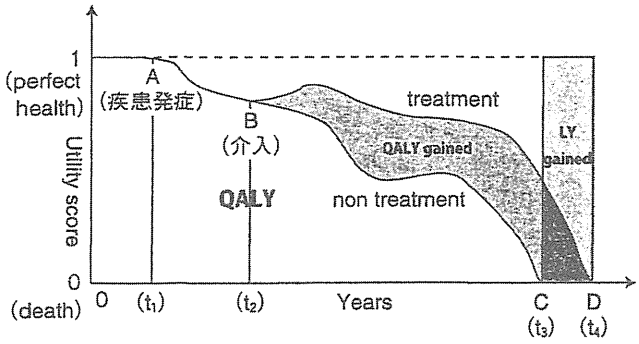

図2　QALY と LY の関係

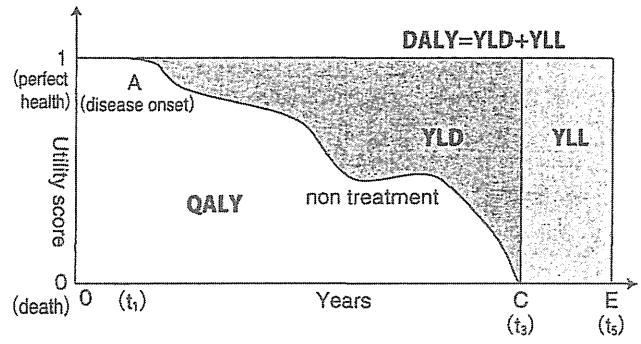

図3　QALY と DALY の関係

合には，効用値は B→C の曲線をたどり，点 C（t_3）で死亡する．薬を使うと効用値が少し改善し，B→D の曲線をたどり，点 D（時間 t_4）で死亡する．薬剤経済評価で重要なのは，効果そのものでなく「効果の差分」である．QALY をアウトカムにとるならば，「QALY gained」と示した B，C，D で囲まれた部分の面積が「効果の差」となる．一方で，生存年数（life years：LY）をアウトカムにとった場合は，効果の差は期待生存年数の差 CD（t_4-t_3）に等しくなる．図形的には，「LY gained」で示した底辺 CD，高さ 1 の長方形の面積で表せる．

QALY と対をなす概念として，障害調整生存年（disability-adjusted life years：DALY）がある．図 3 に示すように，DALY は，疾患による QOL の低下部分（years lost due to disability：YLD）と，疾患による平均余命の減少部分（years of life lost：YLL）の和として表現される．ここで，E（t_5）は世界で最も長命の国の平均寿命が用いられる．DALY は，ある疾患がある社会全体へもたらす負担（burden of disease, 疾病負担）の推計などに用いられる．

アウトカム指標として QALY をとったものを，費用効用分析（cost-utility analysis：CUA）と称する．CEA と CUA で，分析手法は基本的に変わらないため，CUA を CEA の一部とする文献もある．1QALY あたりの ICER がいくらまでなら妥当（たとえば，公的医療制度でカバーするのが可能）かについては定まった基準はないが，米国では 5 万ドル（500 万円，1 ドル＝100 円），英国では 2～3 万ポンド（400～600 万円，1 ポンド＝200 円）程度までなら妥当とされる．

さらに，アウトカム改善効果をも金銭価値に換算して評価し，コスト増分との大小比較を行う分析を費用便益分析（cost-benefit analysis：CBA）とよぶ．たとえば，ある医薬品によって増加するコストが 50 万円で，導入によって 1 人あたり 0.3QALY の QOL 改善効果が得られたと仮定する．費用便益分析では，「1QALY 改善」をさらに金銭価値に換算する．これを便益（benefit）と呼ぶ．1QALY 改善の金銭価値が 500 万円であれば，この薬で得られる便益は 500 万円/1QALY×0.3QALY＝150 万円．コスト増分が 50 万円だから，差し引き 150－50＝100 万円のプラス……と評価する．

CBA は，医療以外の公共事業分野ではよく用いられる手法で，この分析を使えば医療以外の分野との比較，たとえば「医薬品の導入」と「ダムの建設」どちらを優先すべきかなどの評価も可能になる．ただしアウトカムをどのように金銭換算するかは議論も多く，現状では CBA を用いた研究は数少ない．

4.　コストの取り扱い方

薬剤経済評価を含む臨床経済評価におけるコストは，実際にお金が動く「直接コスト」（direct cost）と，実際のお金の動きはない「間接コスト」（indirect cost）とに大別される．ここでは，関節リウマチを例にとって，それぞれにどのようなコストが含まれるかを論ずる．

a. 直接コスト

病態が悪化すれば，当然医療費も増加する．医療費としては薬剤費や検査費，さらに人工関節の手術費など，医療保険でカバーされるものにとどまらず，健康食品や漢方薬などの代替医療のコストも考慮する必要がある．日本の医療用医薬品の市場規模が年間約8兆円なのに対し，代替医療の市場規模も年間3兆円を超えており，疾患領域によっては無視できない大きさになる．

さらに医療費以外のコストとしては，まず介護のコストや通院にかかる交通費，自助具や住宅改造のコストが考えられる．介護や交通費については，患者のみならず，患者の家族にコストがかかることも多い．

b. 間接コスト

間接コストは，働けなくなることによる労働損失を算入するのが一般的である．関節リウマチにかかって働けなくなったことによる労働損失に加え，先ほどの介護や交通費と同様に，家族などの介助者が仕事を休む・仕事を辞めたことによる労働損失もコストに含まれている．

なお，どのようなコストを分析に含めるかという「立場」(perspective) の設定は，薬剤経済評価ではきわめて重要である．たとえば「患者の立場」に立てば，1万円の薬代は通常の自己負担割合30％であれば，1万円×30％＝3000円となる．一方で保険者の立場からは，1万円の薬代のうち負担額は7000円となる．また，通院にかかる交通費は，患者の立場ならば100％算入されるが，保険者の立場ならば算入されない．すべてのコストを広く算入する立場を，社会の立場（societal perspective）と呼ぶ．立場によって推計結果は大きく変動するので，分析の対象に合った立場を選択することが不可欠である．

おわりに：薬剤経済評価の今後

経済評価の結果をどのように解釈するかについては，さまざまな誤解がある．薬剤経済評価の結果は，医薬品の保険償還や採用の可否を一律に定めるものではない．薬剤経済学の目的は，あくまで意思決定の一助となる定量的データを提供することである．「1QALY延長あたり○○万円」という結果が出た際に，実際にその医薬品を使うか否かは，意思決定者に委ねられている．「新薬は高いが，効果もすぐれている．だから使うべき」という定性的な議論でなく，「新薬の導入によってコストは○○円増えるが，平均QALYを既存薬に比べて××QALY増やすことができる．すなわち，あと1QALY増やすのに○○÷××円かかる」という定量的な判断材料を提示することで，より合理的な意思決定をはかるのが薬剤経済評価の目標である．前者のような定性的な考え方のみでは，「新薬はたしかに効果にすぐれているが，コストは高い．ゆえに使うべきでない」という正反対の考え方が出てきた際の反駁が不可能になる．

薬剤経済評価はコストの高い新薬を否定するものではない．むしろ，高額でも著効を示す新薬について，正当な価値付けをするエビデンスを提供するものである．薬剤経済学が，画期的な新薬，また古いが価値の高い薬のより適切な評価につながることを願いたい．　[五十嵐　中・津谷喜一郎]

文　献

1) Drummond, M. F., Sculpher, M. J., Torrance, G. W., et al.: Methods for Economic Evaluation of Health Care Programs, third edition. Oxford University Press, 2005.

2) Gold, M. R., Siegel, J. E., Louise, B., et al.: Cost-effectiveness in Health and Medicine. Oxford University Press, 1996.

3) Shiroiwa, T., Sung, Y. K., Fukuda, T., Lang, H. C., Bae, S. C., Tsutani, K.: International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? *Health Economics*, 19(4): 422–437, 2010.

C15 エンドポイント

臨床試験における評価項目をエンドポイント（endpoint）という．エンドポイントとして定められる項目は臨床試験の目的や疾患などに応じて様々であるが，通常は疾患領域ごとに標準的に用いられるエンドポイントが存在する場合が多い．

糖尿病治療薬の臨床試験を例にすれば，空腹時血糖値や，HbA1C，目標 HbA1C 値に到達する割合，糖尿病性合併症の発生割合，生存期間，health related quality of life（HRQoL）など種々の有効性評価項目が考えうる．また，これらの項目をどの時点で（試験開始から半年後？ 1年後？ 2年後？……）評価するかによっても結果は異なってくる．そのため，試験開始から「どの時点」で「どの項目」を用いるか事前にきちんと定めておき，解析方法などを含めてプロトコルに記載しておく必要がある．

当然のことながら，エンドポイントにとられる項目は科学的に信頼性の高いものでなければならない．かつて日本における治験では全般改善度と称して，医師による主観的な評価が用いられることが多かった．たとえば，治療前後で「著名改善」「改善」「不変」「悪化」「著名悪化」のどれに該当するかを評価することがしばしば行われていたが，そのような曖昧なエンドポイントを用いることは，一部の分野を除いて許容されなくなりつつある．

また，ICH * E9 ガイドライン「臨床試験のための統計的原則」[1] ではエンドポイントを変数（variable）と呼んでいる．

エンドポイントは，その性質に応じて以下のように分類できる．

1. プライマリーエンドポイントとセカンダリーエンドポイント

臨床試験におけるエンドポイントとして様々なものが考えうるが，第Ⅲ相試験のような検証的試験においては，検証の対象となる主要なエンドポイントは一つに絞ることが多い．このような試験の目的であり，検証の対象となるエンドポイントをプライマリーエンドポイント（primary endpoint；主要評価項目）と呼ぶ．

プライマリーエンドポイントを一つに絞るのは，試験全体での第1種の過誤（誤った仮説が検証される確率）を増大させないためである．すなわち，検証の対象となるようなエンドポイントが複数あると，検定の多重性の問題が生じるため，臨床試験全体で誤った仮説が検証されてしまう確率が事前に定めた有意水準を超えてしまう可能性がある．

たとえば，有意水準 5% で個々の仮説を検定するとしよう．また，主要な仮説が5個設定されており，これらの仮説はいずれも正しくないとする．このとき個々の仮説が誤って検証されてしまう確率は 0.05 以下であるが，この5個の仮説のうちいずれかが検証されてしまう確率は最大で，$1-(0.95)^5=0.226$ となる．すなわち，仮説が複数あると試験全体では第1種の過誤が 5% 以下に抑えられなくなる．

一方で検証の対象とはならないが，臨床試験の中で評価したい項目はエンドポイントをセカンダリーエンドポイント（secondary endpoint；副次評価項目）として規定しておくことが多い．たとえば薬剤の有効性を証明したい臨床試験の場合，安全性に関する指標（有害事象の発生率など）はセカンダリーエンドポイントとして評価されるのが一般的である．

2. 真のエンドポイントと代替エンドポイント

高血圧において，血圧が高いこと自体はそれほど大きな問題ではない．高血圧が問題とされるの

はそれが心筋梗塞や狭心症などの心血管系イベントや脳梗塞など，深刻な症状を引き起こすからである．

臨床試験で降圧剤の評価を行うときも，その本来の目的たる心血管系イベントや脳梗塞などをエンドポイントとすることが望ましいだろう．しかし，心血管系イベントの発生頻度は低く，またすぐに起こるわけでもないから，評価には多くの被験者を長期間観察し続けなければならない．結果として，有用な薬剤の承認が遅れることにもなりかねないし，臨床試験に要するコストも課題となる．一方で，降圧剤の血圧への影響は比較的短期間のうちに評価ができる．血圧の大きさがイベントの発生と十分な関係があるならば，臨床試験のエンドポイントを血圧とすることも許容されるだろう．

このとき，心血管系イベントや脳血管系イベントなどの本来測定したいエンドポイントを真のエンドポイント（true endpoint），「臨床的効果を直接測定することが実際的でない場合に，効果の間接的な測定値を示す」[1] 血圧のようなエンドポイントを代替エンドポイント（surrogate endpoint）と呼ぶ．たとえば動脈硬化（心血管系イベント）における LDL コレステロール濃度，糖尿病性合併症に対する HbA1c 値，がんの腫瘍縮小効果と全生存期間（overall survival），痛風の血清尿酸値，HIV における CD4 細胞数などが代替エンドポイントの例としてあげられる．

ただし，実際には設定された代替エンドポイントが真のエンドポイントの予測因子として不適であることが明らかになる場合もある．そのため，第Ⅲ相試験では代替エンドポイントを用いても，市販後の第Ⅳ相試験ではより長期の大規模臨床試験を行い，真のエンドポイントで評価する場合も多い．

代替エンドポイントの統計学的な定義として，Prentice は「代替エンドポイントを用いた治療群間に差がないという帰無仮説の検定が，真のエンドポイントを用いた帰無仮説の妥当な検定となっていること」[2] としている．詳細な議論は文献[3,4]を参照していただきたいが，注意しなければならないことは「代替エンドポイントと真のエンドポイントの相関が十分に大きくても，代替エンドポイントとしては必ずしも妥当ではない」，すなわち相関よりも強い条件が必要になるということである．

たとえば，心血管リスクの高い 2 型糖尿病患者約 1 万人を対象に強化療法群（HbA1c 目標 6%以下）と標準療法群（HbA1c 目標 7.0～7.9%）を比較した ACCORD 試験[5] では，治療後の HbA1c 値は強化療法群で 6.4%，標準療法群で 7.5%であった．しかし，総死亡をエンドポイントとして比較すると，強化療法群が 1.4%，標準療法群が 1.1%と有意に（$p=0.04$）強化療法群の死亡が多くなった．また，HbA1c と総死亡には相関がある，すなわち強化療法群でも血糖コントロール不良群は死亡率が高かったとされている（ここでは，結果の臨床的な妥当性は議論しない）．

真のエンドポイントを総死亡，代替エンドポイントを HbA1c と考えれば，このことは，図 1 のように強化療法が代替エンドポイント（HbA1c）を介さず，なんらかの原因で直接に真のエンドポイント（総死亡）に影響を与えてしまうことを示唆している．そのため，図 2 に単純化して示すように HbA1c と総死亡には相関があるにもかかわらず，HbA1c の低い群の方が逆に死亡率が高くなっていると推測される．もちろん真のエンドポイントと相関があるからといって，この試験において HbA1c を代替エンドポイントとすることは大きな問題をはらむ．

理想的な代替エンドポイントとして，介入の効果が代替エンドポイントを通してのみ真のエンドポイントに影響を与えている図 3 のような状況[6]

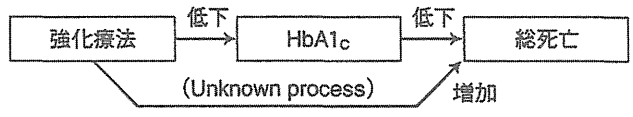

図 1　ACCORD 試験における代替エンドポイントと真のエンドポイントの関係

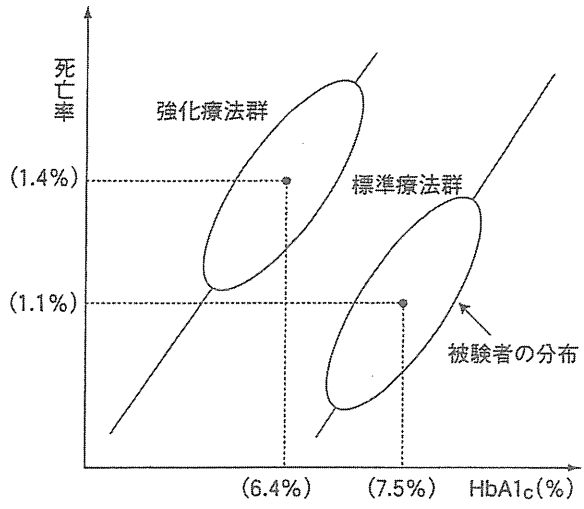

図2　想定される HbA1c と総死亡の関係

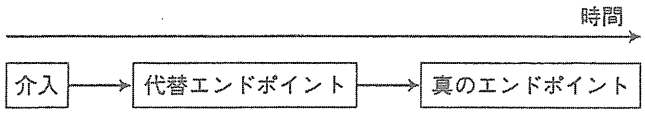

図3　代替エンドポイントが妥当である状況

があげられる．このような状況が成立するには，真のエンドポイントを T，代替エンドポイントを S，治療群を x とし，$f(\cdot)$ を確率密度関数とすると，Prentice の基準によれば

$$f(T \mid S, x) = f(T \mid S)$$

が成り立つ必要がある[2]．これは，介入が真のエンドポイントに与える治療効果は，得られた代替エンドポイントによって完全に表現できることを意味する．もちろん生物学的にこのような条件を満たす代替エンドポイントはきわめて限定されるため，完全な意味での代替エンドポイントを得ることは通常困難であるが，代替エンドポイントの妥当性を検討するための統計学的方法は現在までにいくつかの提案がなされ，応用されている．

3. ハードなエンドポイントとソフトなエンドポイント

臨床検査値や生存時間，イベント発生率，副作用の発現頻度など客観的な評価が可能なエンドポイントをハードなエンドポイントと呼ぶ．一方で，HRQoL や痛み，一部の精神科領域の指標のように評価者の主観によって評価される項目がソフトなエンドポイントである．

とくにプライマリーエンドポイントにおいては，明確な評価が可能なハードなエンドポイントが用いられることが多い．しかし，臨床検査値では数値上の改善が見られても，患者の HRQoL はまったく変わっていないなど，ハードなエンドポイントのみでは必ずしも評価項目として十分でないこともある．そのため，最近の臨床試験ではセカンダリーエンドポイントとして HRQoL などが測定されることも増加してきている．

4. 複合エンドポイント，多重エンドポイント

検証的な臨床試験では，検証すべき項目は一つに絞るのが原則である．しかし，どうしてもプライマリーエンドポイントを一つに絞りきれない場合，ないしは複数の項目を評価することが望ましい場合，以下のような方法が考えられる．

一つは複数の評価項目を単一のエンドポイントに落とし込む方法である．たとえば循環器系の試験では，しばしば複合エンドポイント（composite endpoint）が用いられる．これは様々なイベントをまとめて一つのエンドポイントとして扱うものであり，たとえばプラバスタチンの効果を検証するために日本で行われた大規模臨床試験 MEGA trial[7] ではプライマリーエンドポイントが「冠動脈性心疾患（心筋梗塞含む），狭心症，心疾患による死亡，突然死，冠動脈形成術」と定義されている．複合エンドポイントでは，上記のうちどれか一つでもイベントが起きれば，プライマリーエンドポイントのイベントとして取り扱われる．

また，様々な評価指標を単一のスコアに落とし込むために，合成変数（composite variable）が用いられることもある．たとえば関節リウマチにおける ACR（American College of Rheumatology）コアセットなどがその代表的な例である．

他方で，複数の評価項目をプライマリーエンドポイントとして用いる場合を多重エンドポイント（multiple endpoint）と呼ぶ．通常の多重エンドポイントの検証においては，多重性の調整が必要に

なることが多い．しかし，複数の仮説が積仮説の形になっている場合，すなわちi個のエンドポイント（「エンドポイント1」，「エンドポイント2」，…，「エンドポイントi」）があり，そのすべての仮説を証明しなければプライマリーエンドポイントが検証されたことにならない場合は，第1種の過誤は増加しないため，多重性の調整は不要である．ただし，検出力は低下するため症例数を設計する際には注意が必要となる．また，積仮説でなくても検定の順番をあらかじめ定めておき，仮説の証明に失敗した段階で以降の検定を止めるような階層的な検定手順を用いれば，同様に第1種の過誤を増加させない．　　[白岩　健・津谷喜一郎]

文　献

1) 厚生労働省：臨床試験のための統計的原則ガイドライン，1988.

2) Prentice, R. L. : Surrogate endpoints, in clinical trials: definition and operational criteria. *Stat. Med.*, 8(4): 431–440, 1989.

3) 日本製薬工業協会医薬品評価委員会統計・DM部会：代替エンドポイントの評価，2009.

4) Burzykowski, T., Molenberghs, G., Buyse, M. (eds): The evaluation of surrogate endpoints. Springer, New York, 2005.

5) Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein H. C., Miller M. E., et al.: Effects of intensive glucose lowering in type 2 diabetes. *N. Engl. J. Med.*, 358(24): 2545–2559, 2008.

6) Fleming, T. R., DeMets, D. L.: Surrogate end points in clinical trials: are we being misled? *Ann. Intern. Med.*, 125(7): 605–613, 1996.

7) Nakamura, H., Arakawa, K., Itakura, H., et al.: Primary prevention of cardiovascular disease with pravastatin in Japan (MEGA Study): a prospective randomised controlled trial. *Lancet.*, 368(9542): 1155–1163, 2006.

C16 コンパッショネート使用

「未承認薬」(unlicensed drug) とは，国内においては販売承認されていないが海外では販売承認されている医薬品をはじめ，海外でもまだ承認されていないが有望なデータが得られていて，代替療法のない国内患者に対して有用性が期待される医薬品候補をいう．なお，「未承認薬」と混同されやすい言葉に「無承認薬」がある．無承認薬は，厚生労働省では「無承認無許可医薬品」の用語を使っている．監視取締りの対象となる販売承認も製造許可もされていないのに薬効を謳った偽医薬品のことである．

一方，「医薬品の適応外使用」(off-label use of drug) とは，医薬品は「効能・効果」「用法・用量」などを承認事項として販売承認されるが，医薬品としては販売承認されていても，効能・効果や用法・用量など患者への用い方が承認事項に合致しない場合である[1]．なお，「適応」とはもともと保険適応からきた言葉であり，それが流用されて使われている．

患者にとっての未承認薬使用・適応外薬使用を考える際に，患者の視点からみると，四つの場合がある．

第1は，他に代替治療がない，臨床試験への参加ができない，生命が危険にさらされているなど，とくに重篤な患者の未承認薬へのアクセスである．欧米にはこのような場合にアクセスを例外的に可能とするコンパッショネート使用 (compassionate use of unlicensed drug：CU) といわれる公的な制度[2,3]があり，日本への導入が厚生労働省の検討会で提言されている．コンパッショネートとは「思いやりのある」という意味で，未承認薬の人道的供給を意味する言葉である．一方，日本でこれまで進められていることとして，保険診療で未承認薬使用を患者負担で認める方向がある．

第2は，既承認薬の患者への適応外使用である．ここでは保険で適応外使用が認められるかが問題となる．

第3は，未承認薬・適応外使用の早期販売承認取得による患者アクセスの実現である．エビデンスに基づき新たな臨床試験データを必要としない「公知申請」(後述)，ないしはわずかの追加臨床試験のみでの申請による早期販売承認取得ができる体制の整備が課題である．

第4は，未承認薬・適応外使用の永続的な解消をめざす取り組みである．これについては，皆無とはいかないものの未承認薬・適応外薬を生まないよう，国際共同治験への参加，企業開発へのイ

表1　未承認薬・適応外使用に対して日本でとられた施策 (福澤ほか，2011，一部改変)[4]

稀少疾患，小児科領域および抗がん薬などの領域での未承認薬・適応外使用に対して以下の施策が取られている．

1) 薬事法及び医薬品副作用被害救済・研究振興基金法の一部改正 (「オーファンドラッグ法」) (1993.4) (未)
2) 適応外使用に係わる医療用医薬品の取り扱いについて (「2課長通知，医薬審第104号」) (1999.2.1) (外)
3) 小児用量設定などの試験促進のための再審査期間延長 (2000.12-) (外)
4) 医師主導治験制度の導入 (2003.7) (外・未)
5) 抗がん剤併用療法に関する検討会 (2004.1-2005.2) (外)
6) 未承認薬使用問題検討会議 (2005.1-2009.10) (未・外)
7) 小児薬物療法検討会議 (2006.3-2009.7) (外)
8) 高度医療評価制度の実施 (未承認薬・適応外薬を使用した先端技術を保険医療と併用) (2008.4-) (未・外)
9) 未承認薬・適応外薬の開発支援補正予算 (未承認薬等開発支援事業) (2009.5) (未・外)
10) 新薬創出・未承認薬適応外薬解消等促進加算 (薬価維持特例，2009.12-) (未・外)
11) 医療上の必要性の高い未承認薬・適応外薬検討会議 (2010.2-) (未・外)

(未)：未承認薬関係，(外)：適応外使用関係

ンセンティブ付与，不採算薬の公的資金での開発などが課題となる．これらはドラッグラグやアンメット・メディカルニーズなどの項と重複するので，本項では第1から第3に関連したテーマを扱う．

未承認薬・適応外使用に対してこれまで日本でとられた施策を表1に示した．

1. 未承認薬への患者アクセス

a. 未承認薬のコンパッショネート使用

未承認薬のコンパッショネート使用（CU）は，命を脅かす疾患や強度の衰弱をもたらす疾患などで治療手段が他になく，臨床試験への参加もできない患者に，未承認薬へのアクセスを可能にする公的な制度である．日本には現在この制度がなく，そのような場合に患者・家族・医師がとりうる手立ては，安全管理など問題の多い「個人輸入」しかない．

CUの特徴は，その名が示すように人道的見地からとられる「例外的措置」であることである．

米国では，研究用薬を治療に使用する形でのCUが発展してきた．1938年に連邦食品医薬品法で研究用薬（Investigational New Drug：IND）を規制する権限を得た食品医薬品庁（FDA）は，これらの患者に対する研究用薬の人道的供給に以後一貫して努めてきている．エイズの大流行を受けて1987年には研究用薬の治療使用（Treatment IND）が法制化された．

欧州連合（EU）各国でも，未承認薬へのアクセスの関心は高く，各国またはEUレベルで種々の制度がつくられている．EU各国では米国と異なり，外国では承認されているが国内では承認されていない未承認薬を，必要な患者に輸入して供給する形を主として制度が形づくられてきた．EUの法体系でも上位のRegulation（規則）に「コンパッショネート使用」を位置付け，制度の具体的な運営は加盟各国にゆだねる形がとられている．

韓国においても，未承認薬の人道的供給について制度的な取り組みが進んでいる．1999年に韓国オーファンドラッグセンターが創設，薬事法にも条文化され，外国で承認されている未承認薬を輸入し患者に供給する欧州型のCUを重点に活発な活動を行っている．また2003年には臨床試験承認制度（IND制度）導入実施と併せ，「研究用薬の治療使用」（Treatment IND）制度が導入され，米国型のCUも整備された

日本は，このようなCUの制度をもたないが，未承認薬の人道的供給が公的に行われてきた分野がある．国際交流の活発化で1970年代後半に輸入感染症が増加，治療薬の確保が急務となった．熱意病研究者が当時の厚生省薬務局と協議し，熱帯病治療薬研究班（略称）を発足させ，輸入した医薬品を治験薬の形で無償供給するアクセスルートを開くとともに，関連した研究を推進することが決定された．世界的にCUの法制化をもたらしたエイズも，熱帯病同様対処が急務であった．熱帯病の経験を生かし，1996年にエイズ治療薬研究班（略称）が組織された．これらの研究班は現在も継続されている．他の例として，生活保護での未承認薬に関する医療扶助がある．2008年3月厚生労働省は，同省の検討会において早期の承認が必要と判断された医薬品については，公費供給の対象とすることを認める通達を出した．

ドラッグラグがほとんどない米国でもCUが制度化されているように，CUは保健衛生の観点から社会に不可欠の制度である．日本では2007年7月，厚労省の「有効で安全な医薬品を迅速に供給するための検討会」が，CUの導入に向けて検討を開始すべきと提言した．さらに2010年4月，厚労省の「薬害肝炎事件の検証及び再発防止のための医薬品行政のあり方検討委員会」最終提言が，個人輸入の規制と合わせ，CU等の人道的な例外的使用システムの構築を提言した．後者では，①患者の未承認薬への例外的なアクセスの要望，②患者の安全性確保，③販売承認に必要なエビデンスを「つくる」臨床試験の円滑な実施を妨げないこと，の三つのバランス保持が制度設計の要と指摘されている．このことに留意しつつ，CUの早

期制度化が強く期待される．

b. 「高度医療」による保険医療での未承認薬の患者負担での使用

日本は世界に冠たる国民皆保険の国である．保険外の医薬品使用を含めて「混合診療」は長く論議されてきた．「混合診療」は政治的テーマでもあり，制度は複雑でやや流動的である．2006年10月からは「保険外併用療養」は「評価療養」と「選定療養」に分けられた．前者には，先進医療，治験にかかわる診療，など7種類，後者には差額ベッド，歯科の金合金，など10種類が定められている．

患者からの医薬品アクセスは，このうち「先進医療」での解決がある程度可能である．第2項先進医療は，すでに日本で承認されたもので保険未収載のものである．第3項先進医療は，2008年4月からは「高度医療」とも称され，薬事法上の承認が得られていない未承認薬や適応外使用を用いたものである．高度医療の適合性評価・確認は高度医療評価会議が行い，その先進医療を実施する医療施設を決め，保険診療との併用を認め，承認申請につながるデータ収集の迅速化がはかられている．

2. 既承認薬の「適応外使用」

適応外使用と一口に言っても，その実態は有効性・安全性のエビデンスが十分にあるものからそうでないものまで様々である．本項では有効性・安全性のエビデンスがあり，患者の合意のもとで行われ，患者にとって利益となる適応外使用を念頭に述べる．

適応外使用について保険適用での容認基準を示したものに，1980年（昭和55年）9月の社会保険診療報酬支払基金理事長あて厚生省保険局長通知「保険診療における医薬品の取扱いについて」があり，「55年通知」として知られている．その内容は，有効性および安全性の確認された医薬品（副作用報告義務期間または再審査が終了した医薬品をいう）を薬理作用に基づいて処方した場合は，効能・効果に書かれた病名記載でなくとも認めるものであった．なお，これについて司法は，効能効果と無関係な「薬理作用のみに基づく処方」は認めていないと解釈している（1998年高松高裁判決）．

適応外使用に関する厚生当局の取り組みは，1995年厚生科学特別研究「特定疾患調査研究事業に関わる医薬品の適応外使用に関わる調査研究」に始まる．1996，1997年と類似の厚生省関連研究が続き，1997年の二つ目の研究が，厚生科学研究「難治疾患・稀少疾患に対する医薬品の適応外使用実態のエビデンスに関する調査研究」としてなされた．そこではエビデンスに基づく医療（evidence-based medicine：EBM）の考え方に基づきシステマティック・レビューの方法論が採られ，適応外使用は408件のリサーチクエスチョン（RQ：医薬品と疾患との組み合わせ）として特定され，このうち63件が重要なRQとして選択された[1]．

これにより適応外使用の実態とエビデンスのグレードが明らかになったことで，1999年2月に厚生省から「2課長通知」として知られる医薬審第104号が出され，適応外薬の早期承認への道筋が開かれることになった．これについては次項で述べる．

また，これらの厚生科学研究を引き継ぎ，1998～2000年度にヒューマンサイエンス振興財団の調査研究が実施された．今回は小児科分野を含む日本のすべての適応外使用を対象とした調査で，合計954件のRQが報告され，全RQを含むデータベースが作成された．このうち821件についてエビデンスグレードが評価され，半数を上回る455件（55%）がRCTないしCCT（比較臨床試験）によるエビデンスを有することが確認された[5]．

適応外使用の保険での取り扱いに関する「55年通知」（1980）は現在も生きている．支払基金は再審査が終わった品目について学会等の要望を整理し，認められるものを2007年9月医薬品47事例

(2008年1月1事例削除)，2009年9月医薬品33例，2011年9月医薬品80事例の計159例について55年通知にかかわる情報提供を行っている．しかし，これらの検討を行った情報提供検討委員会の委員名，議事は公開されていないなど，透明性の高いものとなっていない．

3. 未承認薬・適応外薬の早期販売承認取得による患者アクセス

a. 「2課長通知」から未承認薬使用問題検討会議まで

1999年2月に「2課長通知」として知られる厚生省健康政策局研究開発課長と医薬安全局審査管理課長の連名による「適応外使用に係る医療用医薬品の取り扱いについて」(医薬審第104号)が出された．この通知により，米国などですでに承認され，医療における相当の使用実績があり，①それらの承認申請に添付された資料の入手が可能，②信頼できる学術雑誌に掲載された根拠論文または国際機関で評価された総説（引用者注：コクラン・システマティック・レビューのこと）がある場合，ならびに，③国内での公的な研究事業の委託研究で実施されるなど信頼しうる臨床試験成績がある場合は，それらの資料により適応外使用が「医学薬学上公知」であると認められ，新たな臨床試験を行うことなく承認申請が可能になった．公知申請と称される．

厚労省の適応外薬・未承認薬に関する他の施策として，2002年7月の薬事法改正により医師主導による治験が可能となった．これを受け，日本医師会は2003年8月治験推進センターを開設，医師主導治験による効能追加を目指す動きが始まった．医薬品適応外使用のリサーチクエスチョンの調査も開始された．

2004年1月には，社会的に重要性の高いがん治療について，厚労省に「抗がん剤併用療法に関する検討会」が設置された．この分野では世界的に複数の抗がん剤による併用療法が標準治療となっている．しかし日本では抗がん剤としては承認されていても効能が承認されていないなどで，併用療法に用いることが困難な状況にあった．この問題の解決のため検討会が設置され，併用療法に必要な効能の追加が精力的に取り組まれた．61療法がリストアップされ，2005年2月に検討会が終了するまでに21報の検討会報告書が作成され，報告書が取りまとめられた抗がん剤について効能追加がなされた．

辻・津谷は，2課長通知（1999年2月），抗がん剤併用療法検討会（2004年1月～2005年2月）などの取り組みにより，2005年12月までの6年余に56件の適応外使用が正式の承認を取得したことを明らかにしている[6]．

一方，未承認薬については2005年1月，厚労省に「未承認薬使用問題検討会議」が設置された．「混合診療」問題で患者の未承認薬へのアクセスが大きな議論となり，2004年12月厚労相と規制改革相とのあいだで，検討会の設置が合意された．設置目的としては，欧米で承認されているが国内では未承認の医薬品について，①欧米諸国での承認状況および学会・患者要望を定期的に把握，②臨床上の必要性と使用の妥当性を科学的に検証，③当該未承認薬について確実な治験実施につなげることにより，その試用機会の提供と安全確保をはかりつつ，未承認薬の使用を推進するとされた．

検討会議は年4回定期的に開催，また必要に応じ随時開催し，最長でも3カ月以内に結論を出す．欧米で新たに承認された薬は自動的に検討の対象とし，患者の要望に適確に対応．企業治験と医師主導治験に振り分け，確実な治験実施につなげるとした．

検討会議は2009年10月まで22回にわたり開催された．検討品目は44，内訳は抗がん剤22，先天代謝異常症などの小児用薬11，その他11である．2011年9月現在で，44品目中すでに26品目が承認済みであり，国内で治験実施中が15品目，治験計画等検討中が3品目となっている．

2006年3月には「小児薬物療法検討会議」が創

設された．未承認薬については未承認薬使用問題検討会議で扱うこととし，この検討会議では小児に対する適応外使用の実態とエビデンスの把握，それらの結果に基づく対処（企業への添付文書改定や効能追加についての情報提供など）が検討された．検討会議は2009年7月まで6回開催され，アセトアミノフェン・メトトレキサート・A型ボツリヌス毒素の効能効果が承認されるなどの成果があった．

b.「新薬創出・適応外薬解消等促進加算」（医療上の必要性の高い未承認薬・適応外薬検討会議）

2009年6～8月，厚労省は「医療上の必要性が高い未承認の医薬品又は適応の開発の要望についての意見募集」を行った．学会などから374件（未承認薬89件，適応外薬285件）の要望が出され，その評価などは未承認薬使用問題検討会議を改組した新たな有識者会議が，未承認薬・適応外薬両方を扱うことになっていた．

2009年12月，未承認薬・適応外薬への取り組みは注目の新たな段階を迎えることになった．中央社会保険医療協議会（中医協）で，保険薬価算定に関して2010年4月から「新薬創出・適応外薬解消等促進加算」の試行的導入が決まった．この加算は製薬協が求めていた新薬の「薬価維持特例」が形を変え認められたものである．その条件として厚労省が従来の「未承認薬使用問題検討会議」「小児薬物療法検討会議」を統合して新設する「医療上の必要性の高い未承認薬・適応外薬検討会議」の結果を踏まえて開発要請する未承認薬・適応外薬の開発を2年後の改定までに企業が実行することとされた．

この検討会議は2010年2月に発足し，公募で出された要望のうち，109件（未承認薬50件，適応外薬59件）が適応疾病の重篤性，医療上の有用性の観点から開発が必要と判断した．厚労省はこれを受け，5月中旬91件（未承認薬43件，適応外薬48件）について開発権をもつ企業に開発を要請し，国内に該当企業がない17件については開発企業を公募した．

検討会議は，2011年10月現在，9回にわたり開催され，167件の開発を企業に要請（うち42件はすでに承認されている），19件について開発企業を募集している．また，2011年8月，医療上の必要性の高い未承認薬・適用外薬の第2回開発要望を募集した．

薬価加算と開発実行との関係では，要請を受けた企業は6カ月以内の公知申請，ないしは1年以内の治験開始が算定条件となっている．本加算は，従来のオーファンドラッグ開発支援とはインセンティブの方向が逆で，開発に着手しない場合にペナルティを与えるものとなっている．他国にないユニークなものだが，今後の注意深い観察と患者アクセス改善の定量的把握が必要である．

[寺岡章雄・津谷喜一郎]

文　献

1) 津谷喜一郎・清水直容編：医薬品適応外使用のエビデンス．デジタルプレス，1999.

2) 寺岡章雄・津谷喜一郎：未承認薬のコンパッショネート使用―日本において患者のアクセスの願いにどう応えるか．薬理と治療，38(2)：109-150，2010.

3) 寺岡章雄，津谷喜一郎：日本で承認されていない薬を安全に使う―コンパッショネート使用制度―．日本評論社，2011.

4) 福澤　学，井上雅夫，津谷喜一郎：日米における医薬品適応外使用とその施策―1990年代後半以降の歴史・現状・将来―．医薬品医療機器レギュラトリーサイエンス，42(4)：346-356，2011.

5) 津谷喜一郎：医薬品の適応外使用―20世紀末のエビデンス―．ライフサイエンス出版，2004.

6) 辻　香織・津谷喜一郎：エビデンスからみた適応外使用の妥当性．EBMジャーナル，7(3)：8-17，2006.

STUDY

Validation of a New International Quality-of-Life Instrument Specific to Cosmetics and Physical Appearance

BeautyQoL Questionnaire

Ariel Beresniak, MD, MPH, PhD; Yolaine de Linares, MSc; Gerald G. Krueger, MD; Sergio Talarico, MD; Kiichiro Tsutani, MD, PhD; Gérard Duru, PhD; Geneviève Berger, MD, PhD

Objective: To develop a new quality-of-life (QoL) instrument with international validity that specifically assesses cosmetic products and physical appearance.

Design: In the first phase, semidirected interviews involved 309 subjects. In the second stage, an acceptability study was performed on 874 subjects. Thereafter, we recruited a total of 3231 subjects, each of whom completed the BeautyQoL questionnaire, a clinical checklist for the skin, the generic QoL 36-Item Short Form Health Survey, and a sociodemographic questionnaire. A retest was performed 8 days later on a subgroup of 652 subjects.

Setting: Populations in France, the United Kingdom, Germany, Spain, Sweden, Italy, Russia, the United States, Brazil, Japan, India, China, and South Africa, representing 16 languages.

Participants: The general adult healthy population, including women and men.

Main Outcome Measures: Psychometric properties, construct validity, reproducibility, and internal and external consistency.

Results: General acceptability was very good in the 16 languages, with a very low rate of no answers. The validation phase reduced the questionnaire to 42 questions structured in the following 5 dimensions that explained 76.7% of the total variance: social life, self-confidence, mood, energy, and attractiveness. Internal consistency was high (Cronbach α coefficients, 0.93-0.98). Reproducibility at 8 days was satisfactory in all dimensions. Results of external validity testing revealed that BeautyQoL scores correlated significantly with all 36-Item Short Form Health Survey scores except for physical function.

Conclusion: These results demonstrate the validity and reliability of the BeautyQoL questionnaire as the very first international instrument specific to cosmetic products and physical appearance.

Arch Dermatol. 2012;148(11):1275-1282

HEALTH-RELATED QUALITY of life (QoL) is an important clinical objective that has gained prominence during the past few decades.[1,2] Research shows that an improvement in facial attractiveness is associated with positive changes in emotional and social dimensions of one's life, such as personality, interpersonal relationships, and self-esteem.[3,4] Although most clinical research on cosmetic intervention focuses on psychological benefits of cosmetic camouflage, the additional benefits of facial attractiveness achieved with makeup should not be neglected because they may affect someone's life positively.[5,6]

People whose physical appearance has been altered because of a transient or chronic clinical condition, such as pigmentary disorders, are at a higher risk of negative emotional distress due to their altered facial characteristics.[7] A study in China established that patients with vitiligo experienced significantly impaired QoL and unstable marital relationships.[8] Patients with acne have been shown to experience levels of social, psychological, and emotional distress similar to those reported in patients with asthma, epilepsy, and diabetes.[9] To mask their dermatological conditions, these patients may resort to simple makeup application, such as concealing oily skin or acne[10,11]; to mechanical camouflage, such as for vitiligo[7,12,13]; to invasive interventions, such as injectable facial rejuvenations in patients with immunodeficiency syndrome who have lipoatrophy[4]; or to cosmetic surgery. All these studies report significant improvement in patients' well-being after their cosmetic intervention.

Author Affiliations are listed at the end of this article.

Constructing a new QoL instrument requires a systematic conceptual approach and a robust scientific method. Most studies that use QoL questionnaires to assess cosmetic effects were conducted on a limited number of subjects. Moreover, most studies used a narrow validation process specific to the culture of the population being studied. This process can result in findings that are not valid owing to a potential bias from researchers who may select questions from various existing questionnaires, to random effects within a small number of patients, or to a potentially poor interest in a measure that may not be relevant in the country where the study is being conducted. Our objectives were to develop a new QoL instrument that specifically assesses QoL relevant to cosmetic products and physical appearance and to ensure its international validation.

METHODS

POPULATION

The validation of the BeautyQoL questionnaire was conducted among healthy adults in the context of an international multicenter study coordinated by a steering committee composed of 2 senior dermatologists (G.G.K. and S.T.), 1 evaluation expert (K.T.), 1 expert in health-related QoL (A.B.), 1 expert in applied mathematics (G.D.), and 1 expert from the cosmetic industry (Y.D.L.). The subjects were recruited from February 1, 2006, through May 31, 2009, in the following 13 countries representing 16 languages: France, the United Kingdom, Germany, Spain, Sweden, Italy, Russia, United States, Brazil, Japan, India (representing Hindi and English speakers), China, and South Africa (representing Zulu, Sotho, and English speakers).

We included adults (aged 18-78 years) who gave informed consent to participate in the study and spoke the tested language as their native language. The validation of the BeautyQoL questionnaire was conducted on healthy adult subjects in the context of an international multicenter study coordinated by the steering committee described in the previous paragraph. Hence, according to European rules (Directive 2001/20/EC of the European Parliament and of the Council of April 4, 2001; available at http://www.eortc.be/services/doc/clinical-eu-directive-04-april-01.pdf), specific approval from an ethical committee was waived.

STUDY DESIGN

The development of the BeautyQoL questionnaire followed a classic 3-phase validation process using a codevelopment approach by which surveys are conducted in parallel in all participating countries. This approach was favored over a sequential approach by which cross-cultural validation is conducted in sequence, one country after another. Because of the level of management and resources needed for coordinating simultaneous validation in multiple countries, codevelopment approaches are rarely used in the field of QoL assessment. However, this approach was preferred to complete the international validation of this new QoL instrument in a reasonable amount of time.

For the main validation survey, subjects underwent evaluation at inclusion, and 25% underwent retesting 8 days later. The self-administered survey materials that were completed by the patients alone included the tested questionnaire (BeautyQoL) and the validated generic QoL 36-Item Short Form Health Survey (SF-36).[14] The SF-36 consists of 36 items describing the following 8 dimensions: physical functioning, social functioning, role-physical problems, role-emotional problems, mental health, vitality, bodily pain, and general health. Each dimension score ranges from 0 to 100, with a higher score indicating a better perceived state of health. In addition, sociodemographic data and 1 questionnaire regarding skin condition were collected. Subjects underwent retesting on day 8.

DEVELOPMENT OF THE BEAUTYQoL QUESTIONNAIRE

The initial item-generation phase included face-to-face semistructured interviews of 309 subjects. The interviews were conducted by trained clinical psychologists simultaneously in the following 10 countries: France (n=32), the United Kingdom (n=18), Germany (n=46), Spain (n=27), Sweden (n=19), Russia (n=16), the United States (n=53), Brazil (n=32), Japan (n=48), and China (n=18). These interviews aimed at identifying recurrent themes and were used to generate individual questions for the questionnaire. The interviews addressed the effect of cosmetic products and physical appearance on the individuals' QoL. The interviews were also used to determine the wording to be used in question stems and the types and ranges of possible answers. Interviews were conducted in each of the target countries until no new ideas emerged from the content analysis performed in real time. Although local languages were used throughout the interviews, statements were translated to English by a bilingual clinical psychologist and compiled into a standardized interview report for each country. Final semantic content analysis was then performed and complemented by a computerized text-mining analysis (Alceste software; Image, France).

During the second phase of the questionnaire development, an acceptability study was performed on 874 subjects from the original 10 and 3 new countries representing 16 languages, including France, the United Kingdom, Germany, Spain, Sweden, Italy, Russia, the United States, Brazil, Japan, India (representing Hindi and English speakers), China, and South Africa (representing Zulu, Sotho, and English speakers). Subjects were asked to comment on any aspects of the questionnaire (ie, content, wording, and response choices) that they felt were irrelevant or needed improvement. The items that were ambiguous, misunderstood, or rarely answered were excluded or reworded. This acceptability study ensured content validity and guaranteed that the questionnaire was a true reflection of the subjects' perspective in the 16 languages represented. An item reduction was then performed using specific statistical techniques for tracking potential statistical links between questions. We performed κ tests on each question vs all others. A κ:κ value ratio greater than 50% suggested a statistical relation between the questions. Kendall correlations were performed to confirm the κ test results. Correlation coefficients greater than 0.7 suggested a statistical link between the questions. Finally, principal component analyses (PCAs) were used to compare vector distances between questions.

In the third phase, 3231 subjects were recruited in the 13 target countries to fill out the 4 following questionnaires: the BeautyQoL questionnaire, a clinical checklist for the skin (face and body skin characteristics, eg, type, tone, elasticity, and wrinkles, and potential minor problems, eg, spots, scars, broken veins, and being subject to sun reactions or allergies), the SF-36 questionnaire, and a short sociodemographic questionnaire. Subjects were selected from existing general population panels according to socioeconomic criteria specific to each participating country and managed by local survey agencies. A retest of the BeautyQoL questionnaire only was performed 8 days later on a subgroup of 652 subjects (about 40 subjects per target language). The whole database, including all answers pro-

Table 1. Sociodemographic Description of the Study Population at the Validation Phase

	Subjects, %															
	UK	Swe	Jpn	Itl	Brz	Chn	US	Frn	Ger	Ind/ Eng	Ind/ Hindi	Rus	SA/ Eng	SA/ Sotho	SA/ Zulu	Spn
Sex																
Male	50.0	50.0	50.0	50.0	50.2	50.0	50.0	50.0	50.0	48.5	52.9	50.0	50.5	50.2	50.7	50.0
Female	50.0	50.0	50.0	50.0	49.8	50.0	50.0	50.0	50.0	51.5	47.1	50.0	49.5	49.8	49.3	50.0
Family status																
Couple	51.0	58.5	34.0	53.5	53.1	65.0	57.0	64.0	66.5	57.3	72.3	72.5	47.5	33.8	21.7	59.0
Single	34.0	27.5	24.5	23.5	5.2	25.0	28.0	31.5	25.0	33.5	15.5	10.0	37.1	49.8	56.2	27.0
Live with relatives	15.0	14.0	41.5	23.0	41.8	10.0	15.0	4.5	8.5	9.2	12.1	17.5	15.3	16.4	22.2	14.0
Educational level																
Primary[a]	5.0	9.0	0	3.0	0	0	0	3.0	2.0	0	0	0	1.5	4.5	1.5	7.0
Secondary[b]	32.0	34.0	32.5	24.0	5.0	10.0	23.0	11.5	45.0	4.4	14.1	12.0	60.4	63.7	68.0	20.5
Tertiary[c]	67.5	57.0	67.5	73.0	99.5	90.0	77.0	85.5	53.0	95.6	85.9	88.0	38.1	31.8	30.5	72.5
Labor status																
Employed	52.5	60.5	62.5	45.5	61.5	88.0	61.0	50.0	61.0	59.2	56.3	87.0	87.1	64.7	73.9	54.5
Unemployed	12.5	15.5	9.0	20.0	17.4	0	19.0	12.5	8.0	2.9	1.5	4.0	3.0	20.9	14.8	22.0
Retired/pensioner	12.0	13.5	5.0	14.0	4.2	2.5	9.5	14.5	14.0	0.5	0.5	5.5	1.0	0	0	13.5
Student	10.5	9.5	3.0	12.0	6.1	9.5	1.5	6.0	6.5	18.4	10.7	2.0	5.4	13.9	10.8	2.5
Homemaker	12.5	1.0	20.5	8.5	10.8	0	9.0	17.0	10.5	18.9	31.1	1.5	3.5	0.5	0.5	7.5
Housing																
Own home	47.0	51.5	51.5	46.5	65.3	63.5	60.5	47.0	29.5	90.8	88.8	78.0	36.1	41.3	29.6	64.0
Renting	36.5	41.0	35.5	21.5	23.9	11.5	28.5	47.5	65.5	9.2	11.2	7.0	39.6	24.9	23.6	18.0
With family	14.5	5.5	9.5	26.0	2.3	18.5	10.0	0	2.0	0	0	9.5	22.8	30.3	42.4	2.5
Other	2.0	2.0	3.5	6.0	8.5	6.5	1.0	5.5	3.0	0	0	5.5	1.5	3.5	4.4	15.5
Primary residence																
Urban	68.5	65.0	100.0	74.0	100.0	100.0	56.5	61.5	62.5	100.0	99.5	100.0	99.5	99.5	98.0	81.0
Rural	31.5	35.0	0	26.0	0	0	43.5	38.5	37.5	0	0.5	0	0.5	0.5	2.0	19.0

Abbreviations: Brz, Brazil; Chn, China; Frn, France; Ger, Germany; Ind/Eng, English-speaking India; Ind/Hindi, Hindi-speaking India; Itl, Italy; Jpn, Japan; Rus, Russia; SA/Eng, English-speaking South Africa; SA/Sotho, Sotho-speaking South Africa; SA/Zulu, Zulu-speaking South Africa; Spn, Spain; Swe, Sweden; UK, United Kingdom; US, United States.

[a] Usually begins at 5, 6, or 7 years of age and lasts for 4 to 6 years.

[b] Generally continues the basic programs of the primary level, although teaching is typically more subject focused.

[c] Largely theory based and designed to provide sufficient qualifications for entry to advanced research programs and professions with a high skill requirement.

vided in a total of 16 languages, was split randomly in 2 subsamples. For 1 subsample, the multidimensional structure of the questionnaire was identified studying interitem, itemdimension, and interdimensional correlations (Pearson correlation tests) and PCA.[15] Varimax rotations applied to the PCA identified the main axes composed by a subgroup of questions, in which the component for 1 particular axis would be greater than 0.5[15] (each subgroup of questions representing 1 dimension). For each potential dimension scale, internal consistency reliability was assessed by the Cronbach α coefficient. A Cronbach α coefficient of at least 0.70 was expected for each scale.[16] Within each dimension, the items for which deletion would lead to an increase in the α value of at least 0.02 were candidates for deletion. The unidimensionality of each dimension was assessed using Rasch analyses. For the second subsample, confirmatory factor analysis[17] was used to assist selection by testing the various candidate scale structures according to different potential item selection patterns. The more meaningful and psychometrically sound construct was kept to produce the final version of the BeautyQoL questionnaire.

To explore external validity, relationships were investigated between the dimensions of the BeautyQoL instrument and the generic SF-36 instrument. The underlying assumption was that dimension scores of the BeautyQoL questionnaire would be more strongly correlated with scores on similar dimensions from the other instruments than with dissimilar ones (convergent validity). The discriminant validity of the BeautyQoL questionnaire was determined by dimension mean scores across subject groups that were expected to differ in their sociodemographic features (eg, age and sex) or clinical features (eg, skin status) using analysis of variance, the Mann-Whitney test, and Spearman correlation tests. Reproducibility was analyzed through test-retest reliability using intraclass correlation coefficients between the 2 successive assessments of subjects selected for retesting.[18] Sensitivity of the BeautyQoL questionnaire was assessed in the frame of a randomized clinical trial conducted in France, which compared the impact of the following 2 camouflage products: cream (high-coverage foundation) vs powder (high-coverage loose powder) used for 3 weeks by 88 subjects with facial cosmetic imperfections.

RESULTS

DEMOGRAPHICS OF THE POPULATION

Among the 13 countries representing 16 languages, we found no significant difference in the sex status. Significant differences were found among the countries in family status (>72% were living as part of a couple in Russia and Hindi-speaking India vs 34.0% in Japan or 21.7% in Zulu-speaking South Africa), educational level (99.5% had a tertiary educational level in Brazil vs <40% in South Africa), labor status (88.0% were employed in China vs <60% in France, Spain, India, and the United Kingdom), housing (about 90% lived in their own home in India vs 29.5% in Germany), and primary residence (≥98% of the population lived in urban areas in Japan,

Table 2. Item Correlation With 5 Dimensions of the International English Version 3.0 of the BeautyQoL Questionnaire[a]

Dimension	Cronbach α Coefficient	Question
Social life	0.98	Have you felt an improvement in your social life?
		Have you felt less sad?
		Have you felt an improvement in your family life?
		Have you felt an improvement of your credibility?
		Have you felt more secure?
		Have you felt an improvement in how people respect you?
		Have you felt an improvement of your social status?
		Have you felt an improvement of your mood?
		Have you felt that people are more willing to trust you?
		Have you felt transformed?
		Have you felt more fun to be with?
		Have you felt an improvement in how you express yourself?
		Have you felt an improvement of your emotional sensitivity?
		Have you felt an improvement of your ability to stay awake?
		Have you felt an improvement of your daily quality of life?
		Have you felt more successful?
Self-confidence	0.97	Have you felt good?
		Have you felt an improvement in your psychological life?
		Have you felt an improvement of your self-esteem?
		Have you felt an improvement in your physical appearance?
		Have you felt more confident?
		Have you felt more pleasure?
		Have you felt more overall satisfaction?
		Have you felt an improvement of your happiness?
		Have you felt an improvement of your sensuality?
Mood	0.96	Have you felt more relaxed?
		Have you felt less stressed?
		Have you felt more your joy?
		Have you felt more motivated?
		Have you felt calmer?
		Have you felt less depressed?
		Have you been satisfied with your actions?
		Have you felt more mobile?
Energy	0.93	Have you felt healthier?
		Have you felt more energetic?
		Have you felt less tired?
		Have you felt an improvement in your physical activity?
Attractiveness	0.93	Have you felt more seductive?
		Have you felt invigorated?
		Have you felt that people pay more attention to you?
		Have you felt an improvement in your vitality?
		Have you felt that you look younger?

Abbreviation: QoL, quality of life.

[a]Specific versions have been developed in 16 languages from the following 13 countries: the United Kingdom, Sweden, Japan, Italy, Brazil, China (in Mandarin), the United States, France, Germany, India (in Indian English and Hindi), Russia, South Africa (in South African English, Sotho, and Zulu), and Spain. Answer modalities include "completely," "a great deal," "somewhat," "not much," "not at all," and "it is worse." From the item-generation phase, 61 items were selected leading to 61 questions in the first prototype questionnaire describing major quality-of-life domains, such as well-being, self-esteem, social life, love life, professional life, sexual life, confidence, happiness, image, status, and emotion. The first item-reduction analysis led to the second prototype composed of 48 questions, then 42 questions after the second item reduction (version 3.0).

Brazil, China, English-speaking India, Russia, and South Africa vs <70% in the United Kingdom, Sweden, the United States, France, and Germany) (**Table 1**).

VALIDATION PROPERTIES

Five axes were identified by the PCA, representing the following 5 dimensions explaining 76.7% of the total variance: social life, self-confidence, mood, energy, and attractiveness. Internal consistency was high (Cronbach α coefficients, 0.93-0.98). Reproducibility at 8 days was satisfactory in all dimensions. External validity testing revealed weak significance for the correlation of BeautyQoL scores with all SF-36 dimensions except for the physical function dimension, which was expected because of the poor link between physical function and appearance (Pearson correlation coefficient, −0.02). These results suggest that BeautyQoL dimensions would capture specific perceptions not covered by generic instruments such as the SF-36.

The second item-reduction analysis led to the final version of the BeautyQoL questionnaire consisting of 42 questions. The international English version 3.0 is presented in **Table 2** with the items correlated with the 5 dimensions. Finally, the clinical trial conducted in France confirmed the ability of the BeautyQoL questionnaire to discriminate subclinical changes in facial cosmetic imperfections. The QoL global score for the group receiving the