

第1章

QI作成の 基本的考え方 と方法



診療の質指標 (Quality Indicator) 作成の基本的考え方と方法

分担研究者

東 尚弘

国立がんセンターがん予防・検診研究センター

1. 診療の質とは何か

「診療の質」という言葉は非常に多くの場で使われる言葉でありながら、その意味が明確に定義されることはあまりない。「診療の質とは、個人および集団に対する診療行為が望まれた健康状態をもたらす確率をあげ、かつ、最新の専門知識と合致する度合いをいう」といった米国医学研究所 (Institute of Medicine) の定義は頻用され公式的なものであるが、もう少し平たくいえば、「診療の質」とは的確なタイミングで適切な診療行為が行われる医療を意味する。医療事故や医療における安全の問題は、診療の質に関する極端な例であるが、最新のエビデンスに基づく治療、過不足のない治療・検査など、診療の質にはいろいろな側面がある。また実際の診断や治療の内容という専門技術的な質だけではなく、患者の満足度に代表されるような対人関係的な質や、快適さを中心に考えた質も考えられる。

本研究では、専門技術的な診療の質に絞った評価のための指標を設定した。昨今、診療の質について世界的な問題となっているのは、最新の臨床知識が現場で生かされていない、普及していない (Evidence-practice gap) ということである。つまり、臨床知識的には標準と考えられるものが、実際の標準となっていないということである。この問題意識を基本として本研究の診療の質評価測定は考えられている。

2. 診療の質指標 (QI) の用語について

本研究班では一貫して質指標 (クオリティインディケーター、Quality Indicator : QI) という言葉を使用している。似たような用語に、臨床指標 (クリニカルインディケーター、Clinical Indicator) やパフォーマンス指標 (パフォーマンスインディケーター、Performance Indicator) といったものがある。それぞれ方言のようなもので、そこに本質的な違いはあまりない。ただ、臨床指標といった場合には、医療費や在院日数などのような経営指標に対比して指標が「臨床」事項として意識されており、パフォーマンスと呼んだ場合には、「質」という言葉を避ける意識が働いている印象がある。さらに、測定を強調して質やパフォーマンスの尺度 (Quality measure、Performance measure) という場合もある。

3. QIの基本的な形と実施率の計算

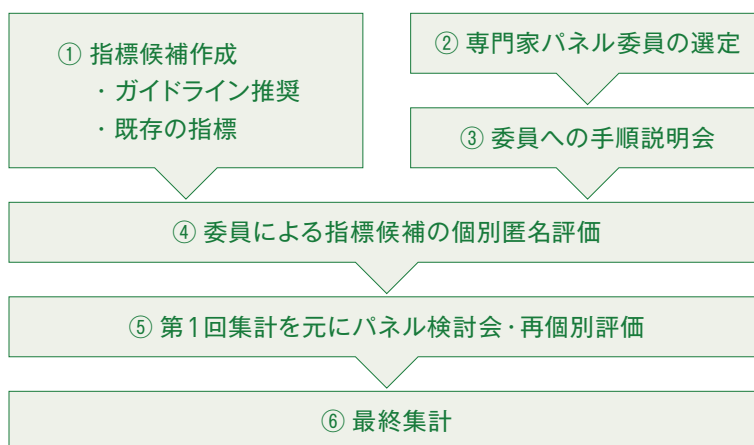
個別のQIは各章をご参照いただきたいが、基本的には分母に対象とする患者や臨床状況を記述し、分子には分母に示されるような患者に行われることが標準と考えられる診療内容が示されている。これら分母の患者数のうち、分子の診療内容を満たす患者数でそのQIの実施率を計算する。例えば、大腸癌の最初のQIでは、手術、化学療法、または放射線治療を受けた大腸患者数が分母であり、治療前に血清CEA値を測定した結果の診療録記載がある患者数が分子になる。QIそれぞれについてこのように計算した実施率を他施設の同一QIと比較、また同一施設の中で別のQIと比較することにより、自施設の診療状況が明らかになり、実施率の低い部分について理由を検討することで質の改善につながる。

4. 本研究におけるQIの作成方法

a. 全体の流れ

今回のQI作成には国際標準とされているQI作成手法を用いた(図1)。この手法は、米国カリフォルニア大学ロサンゼルス校(UCLA)およびランド研究所(RAND Corporation)で開発されたことから、RAND/UCLA 適切性評価法と呼ばれ、別名デルファイ変法と呼ばれることもある。

図1



まず、専門家数名がQIの候補と、候補としての根拠についてエビデンスに基づいたレポートを作成。同時に各臓器のがん診療の専門家10名程度の専門家パネルを設置する。それぞれのQI候補およびまとめられた根拠は、専門家パネルの各委員に郵送され、委員は各自QI候補を吟味、適切性について評価する。続いて、1日程度の検討会議を開催して評価結果の集計を参照しながら検討し、会議の内容を受けてQIについて第2回目の評価を実施する。委員の多くが適切と評価し、かつ反対者が少数であったものだけを最終的なQIとして採用した。このような手順で、それぞれのQI候補について詳細かつ多角的に吟味されたものをQIとして完成する。QIが完成するまでの詳細は以下に述べるとおりである。

b. QI候補の作成

QIの候補となる主な情報源は3つある。1つ目はすでに海外で作成されたQIであるが、技術水準や医療制度の異なる海外の標準がそのまま日本で使えるわけではないので候補として扱うにとどめた。個別には、すでに発表されている、米国臨床腫瘍学会 (American Society of Clinical Oncology) の行ったNICCQ(National Initiative for Cancer Care Quality) プロジェクト、米ジョージア州が州として診療の質を改善するために設定したQI、専門誌の論文、高リスク高齢者医療の質を測定するための第3次ACOVE(Assessing Care of Vulnerable Elders) プロジェクトQI ASSIST(Addressing Symptoms Side Effects and Indicators for Supportive Treatment)を参考とし、日本の状況に合うような形で調整したものを候補とした。2つ目の情報源は、各臓器について専門学会が発行する診療・治療ガイドラインである。ガイドラインの推奨度の高いものの中から、例外が多いと考えられるものや解釈の幅が大きいため基準としては使いにくいものを除外、変更してQI候補とした。これら2つの情報源を使用したものについては、各QIの解説の項に参照元の先行研究やガイドライン推奨が記載されている。3つ目の情報源は専門家の考える新しいアイデアである。手術記録への記載を促した項目や、術前検査で注目すべき画像所見などの提案といったアイデアからQI候補としたものが数多くある。

このようにして定められた候補について根拠をまとめたレポートを作成した。すべてのQI候補に根拠となるランダム化比較試験が存在するわけではないが、疫学データをはじめとする客観的データを総合して、ベスト・エビデンスに基づいて根拠を要約した。ガイドラインや先行研究をもとに作成したものでは、その作成時点までの文献検索はすでに行われているため、作成時以降の文献を中心に検索を行い、最新の知見を盛り込んだ。この作業は各臓器2～5名での分担作業で行った。

c. 専門家パネル委員の選定

各臓器において、9～12名の専門家パネル委員を選定。各臓器の診療に携わる外科医、化学療法や内視鏡の専門家など広い範囲の専門家、診療科を含めた構成とし、大学病院やがん専門施設だけでなく、一般病院の臨床医も意識的に委員として選出した。選出には、診療パターンや地域特性、また可能な範囲での世代的な多様性の確保に努めた。さらに、分野によっては関連する他分野の専門家も含めた。選出された委員が今回の手順を理解するための説明会を学会等に合わせて開催し、説明会欠席者には、個別に説明の機会を設けた。

d. 専門家パネル委員による事前のQI候補評価

選定されたQI候補に図2のような1～9の評価スケールを付けて、パネル検討会の約3週間前に、「根拠のまとめ」とともに専門家パネル委員に郵送し、各委員はエビデンスや自らの判断をもとに、QI候補がQIとして適切かどうか、極めて不適切から極めて適切までの9段階スケールで評価した。評価にあたり、QI候補の表現や内容について別案がある場合は、コメントとして記入するように依頼した。QIとしての適切性は、「原則としてそのような診療を行うべきかといえるかどうか」、また「行っていないことが質の低さに通じるかどうか」、ということを経験している。それを前提に「準拠している割合が多いほど質の高い診療を提供している」といえると考えた。評価後、各委員は評価結果を返送し、事務局が集計を行った。

図2

QI候補	QIとしての適切性	コメント
手術を受ける患者には、手術の利益と不利益の両方が説明される	1 2 3 4 5 6 7 8 9	
がんと診断された患者は疼痛に関して評価がなされ診療録に記載される	1 2 3 4 5 6 7 8 9	
手術以外の治療が行われる患者は治療開始前に確定診断がなされる	1 2 3 4 5 6 7 8 9	

注) QI候補の内容は例であり、実際のものとは異なります。

e. パネル検討会と第2回評価

集計をもとに各臓器について、全員参加のパネル検討会を開催した。ここでは、評価内容と評価した委員の関係には触れず、QI候補をスクリーンに映し出し、内容の修正を加えながら検討した。各QIの協議が終わるごとに、各委員は第2回目の評価を実施、提出した。その中でQIの使われ方等を再度確認し、また根拠についても適宜吟味した。また、議論の過程において新しいQIが提案された場合には、そのQIについて多角的に評価して第2回評価に加えた。

f. 最終集計と採用基準

第2回評価をもとに、原則として中央値が7以上であり、評価スケールで1～3を付けた評価委員数が2名以下のものを採用、最終的なQIセットとした。採用の内容が両立しないもの、例えば、手術後の画像検査によるフォローの「間隔」のみが「3ヵ月おき」というものと「6ヵ月おき」というものが両方とも採用となった場合には、反対の少ないもの、中央値の高いものを採用した。同様の分布の場合には、基準として緩やかなほうを採用した。第2回集計結果を議事録とともに電子メールで回覧し、合意を得て最終確認とした。

g. 本方法の特徴

本研究で採用した手法は、以上のようなものである。この方法には、

- (1) 検討会前に各委員が個別に検討する期間を提供し、
- (2) 適切性を数値で評価することにより各委員の意見を可視化、
- (3) 関連する診療科や地域特性などを踏まえての専門家パネル委員の構成の工夫により多角的検討を実現し、
- (4) 検討会の議論で、QIの解釈などにさまざまな見方が反映された評価方法となり、その結果、完成度の高いQIの作成を可能にした。

また、

- (5) 2回目も個別評価することにより、各委員は検討会の議論の大勢にとらわれずに自らの意見で評価できる

といった特色を包含している。検討会ではかなり白熱した議論が戦わされ、対象患者を絞ったり広げたりするという調整や、新しいQIの追加、修正が随所に加えられた。結果として各分野で策定されたQIは全部で206に上った。各分野、各QIの詳細についてはそれぞれの章を参照されたい。

5. 留意点

これらQIの作成過程で認められたいくつかの留意点を以下にまとめる。

a. 診療としては行われたが記録されていないものをどう扱うか

標準診療がどの程度行われているのかを吟味するために診療録から情報を収集しているが、診療として行われていても、記録されていない場合をどう取り扱うかという問題がある。これに関しては、記載自体が標準診療として考えられるかどうかを基準として、QIの適切性を吟味した。つまり、診療を行うだけで診療録に残ってなくても構わないものに関しては、QIとしては厳格すぎるという意味で不適切と評価することとした。このことは、最終的に採択されたQIの内容が行われていても記録のないものは診療の質は低いと考える、という専門家パネル委員の判断を表している。

b. ガイドライン推奨とQIの違い

ガイドラインとQIは標準診療を扱っているという点で類似している。しかしながら、ガイドラインの目的は最新の治験に関する情報の普及による診療の支援であり、QIの目的は診療の評価である。このことから、いくつか意識すべき相違点がある。第一に対象とする範囲が異なる。ガイドラインは情報の普及によって診療を支援するという立場から、コンセンサスがでない分野について「コンセンサスがでない」ということや、「こうすることで有効な可能性がある」という程度の情報であっても提供することに意味がある。一方QIは、標準とのコンセンサスがでない事柄はQIにはなり得ないため、カバーできない。しいて作るならそのような診療を行う場合には「標準でない旨を患者に説明する」ことがQIとなる。このような工夫により、適応拡大などが積極的に行われている分野における治療などの先進的な試みを妨げることのないように配慮した。

また、ガイドラインの推奨は情報の伝搬・普及が主目的であり、推奨自体は簡潔にし、細部は使用する医師の判断に委ねることも多い。他方、QIでは評価が目的であり、簡潔であるよりは、解釈に幅がないことを旨とする。例えば、手術不能の場合には〇〇を行うのがよいといった推奨は、「手術不能」かどうかの判断は各医師が患者ごとに判断することになるが、QIではできるだけ採録者の判断を入れないで評価できるように、手術が診断から〇〇の期間行われない場合を対象とする、といったように、明確な事実をもとに対象を定義する構造とする。

また「〇〇のような患者に対しては□□が有用である」といった推奨に関しては、QIに変換する際、それが果たして標準であり、原則として□□を行うことが求められるのかといったことを検討して、そうでない場合にはQIとして当該推奨は使えない、という判断を下すことになる。

また、ガイドラインを通じて情報を普及する必要がないような事柄、つまり行った方がよいことは周知であるが、実際には行われていないようなこと、特に診療を円滑にし、それを通じて患者アウトカムを改善するような事柄についてはQIとなる。例えば手術所見・病理所見の適切な記載などはガイドラインには含まれることは少ないものの、情報が見つからないという理由で診療が阻害されることも考えられるところから、QIの内容に含めた。

このように細かい点は異なるものの、ガイドライン推奨とQIは表裏一体の関係にある。両者は同一視するべきではなく、解釈には十分な注意が必要である。

c. 保険の支払いを記載されていない診療を標準としてよいか

本来、標準診療がまず存在し、それらをもとに保険による支払いとして記載されるべきであり、保険により標準診療が影響されてはならない。従って、QIの設定は保険支払いを無視して考えるのが理想的である。しかし、複雑な問題を避けるために、候補作成当初より、保険記載されていないものについてQIとすることは今回は原則として避けるようにした。

d. 各QIの重み付けと、総合点計算について

これまで述べてきた手法により、全臓器で206のQIを策定した。それぞれが、対象患者を分母として定義され、それに対して行われる標準診療を分子として実施率を計算する構造となっている。実施率はそれぞれのQIについて計算され、それらを検討することで自施設の改善点などが明らかになるが、総合点は必要かどうかという別の問題がある。総合点を計算するためには、QIのそれぞれに重要性を加味した重み付けが必要である。しかし今回のQI作成にあたり、内容の指標としての適切性に焦点を絞ったため、それぞれの相対的重要性については検討していない。また、本来QIの重要性は、アウトカムを改善することにどのくらい寄与できるかということによって決められるべきであるが、治療行為をそのまま対象としたQIばかりではなく、診断過程の適切性やフォローアップなどのアウトカムに影響する要素が異なるQIも多い。例えば、診断過程は、それに基づいて続いて行われる治療が適切であって初めてアウトカム改善に結びつくため、一律横に並べて直接的なアウトカムとの関連の強さをみるのは困難である。これらの問題点から、世界的にもこれらのQIの総合点の算出方法について確立したものは存在しないのが現状である。

一方、うまく総合点が算出できたとして、その総合点が何を意味するのか、どう改善に結びつくのかということが逆に不明瞭になってしまう一面もある。個々のQIを吟味していると、改善点はそのQIの指定する診療行為にあることは明白であり、それが過程評価のQIの長所でもあった改善点を直接指し示すということを弱める結果になる。多施設との総合的な「質の比較」が目的であれば総合点の算出は必須でありながら、改善という意味においては弱くなるという矛盾をはらんでいる。

以上にかんがみて今回は特に総合点を算出するための試みは行っていない。これらは今後の研究課題といえよう。

e. カバーする範囲について

QIの検討会でQIを追加することはたびたびあったが、採択後も全く新しい事柄で「このようなQIはないのか」という意見がないわけではない。実際に大腸癌・胃癌・乳癌では術前の抗生物質の使い方のQIがあるのに対して、肺癌、肝癌ではそれが存在しない。これらの範囲について、明確な切り分けは行われていない。今後QIを発展させる際に検討すべきと考えられる。

6. 実際の測定にあたって

実測にあたって、筆者が考える標準的手法について以下に述べる。まず、対象である。基本的に院内がん登録等の情報源から、まず患者リストを作成し、その患者の診療録を検討することでQIへの実施率を調べる。その際、最初から標準診療が行われることが見込めない例につ

いては除外してよいと考えている。目的によって使い分け、自施設の質を検証するためには、すべての症例を対象として評価した後に理由を考えるという方法ももちろんあり得るが、もし、多施設比較を行って改善の推進力にしようとするならば、あらかじめ対象症例の範囲を明確にしたデータを収集することが必須になる。以下はわれわれが、QIの予備的研究を行った際に基本的に除外条件と考えたものである。

- ・同時重複癌があるもの（治療過程が変わるため）
- ・介入を伴う臨床研究に入ったもの
- ・原発不明癌など診断が治療前に不明であり、後から病理所見などの予想しなかった最終診断がついたもの。（例えば卵巣癌と思って手術をしたら胃癌だった、など）
- ・他院での治療途中で転院してきたもの
- ・組織型が典型的な癌ではないもの（肉腫・リンパ腫など）

このように、比較をする場合にはその対象を一定の基準にそろえることが必須であることを意識することが必要である。しかし、あまりに最初から対象を絞ると分母が小さくなる。そのため「比較」はしても「競争」ではないことを意識して、実際の結果から多角的な検討を加えることに重点を置き、できるだけ多くの情報をとらえることが必要と思われる。

また、採録の方法であるが、それぞれの臓器の中で複数のQIについて共通の項目も多いことから、情報収集フォームを使用することが効率的な採録につながると考えている。具体的な方法については、研究班事務局で、問い合わせに応じて情報提供を行う予定としている。

参考文献

1. The Institute of Medicine. Medicare: A Strategy for Quality Assurance. Washington D.C: National Academy Press; 1990.
2. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, van het Loo M, McDonnell J, Vader JP, Kahn KL. The RAND/UCLA Appropriateness Method User's Manual: RAND; 2001.
3. Malin JL, Schneider EC, Epstein AM, Adams J, Emanuel EJ, Kahn KL. Results of the National Initiative for Cancer Care Quality: how can we improve the quality of cancer care in the United States? J Clin Oncol 2006;24:626-34.
4. The Institute of Medicine. Assessing the Quality of Cancer Care: An Approach to Measurement in Georgia. Washington D.C: National Academy Press; 2005.
5. McGory ML, Shekelle PG, Ko CY. Development of quality indicators for patients undergoing colorectal cancer surgery. J Natl Cancer Inst 2006;98:1623-33.
6. Wenger NS, Roth CP, Shekelle P. Introduction to the assessing care of vulnerable elders-3 quality indicator measurement set. J Am Geriatr Soc 2007;55 Suppl 2:S247-52.
7. Dy SM, Asch SM, Naeim A, Sanati H, Walling A, Lorenz KA. Evidence-based standards for cancer pain management. J Clin Oncol 2008;26:3879-85.
8. Walling A, Lorenz KA, Dy SM, Naeim A, Sanati H, Asch SM, Wenger NS. Evidence-based recommendations for information and care planning in cancer care. J Clin Oncol 2008;26:3896-902.