

費用対効果の導入に向けた課題

早稲田大学教授 野口 晴子

試行的検証結果の乖離

2017年11月10日、中医協の費用対効果評価・薬備・保険材料の各専門部会の合同部会において、費用対効果評価の試行的導入に際し実施された企業による分析と、有識者を含む第三者による再分析の結果が大きく異なる品目があるという報告がなされた。合同部会においては、対象集団や介入方法、比較対象品目(技術)等の分析の前提要件、そして、分析に用いるデータの選択基準が異なっていることが、結果が乖離した原因であるという見解が示された。

要件を擦り合わせることは重要なことであるが、それ以上に考慮されるべきは、分析に用いるデータや手法の違いである。医学におけるゴールデンスタンダードである実験型のランダム化比較試験(RCT)と、最新のNDB等を用いた非実験型による観察研究にはそれぞれメリットとデメリットがあることが知られており、費用対効果分析を行う場合には、それらを十分念頭に置く必要がある。統計的な信頼性が高いと言われていたRCTを実施するためには莫大な費用がかかるうえに、高齢者や重篤な患者等、リスクの高い人々が実験から除外される可能性が高く、倫理的にも問題が発



生しやすい。したがって、RCTと比べると、統計上さまざまな問題が生ずる可能性は否めないが、どうしても観察研究による補完が必要となってくる。

なぜ結果が乖離するのか?

「RCT」対「観察研究」

情報がRCTによらず観察研究である以上、分析の結果は、たとえば、年齢、性別、重症度、副疾患の有無、家族構成、地理的環境、経済状況、保険種別等、個々の患者属性に大きく依存するため、治療や政策の影響を正確に推定することは、統計学の分析上、困難を極める。このように、さまざまな属性が影響して分析結果を偏らせるという問題が、いわゆる「選択バイアス」と呼ばれる現象である。ここでは、一事例として、1998年のNew England Journal of Medicine誌に掲載された米国におけるRCTと(旧)米国保険財務庁と米国医学会の呼びかけにより収集された観察研究(Collaborative Cardiovascular Project: CCP)の違いを比較し

てみよう(詳細については、拙論、日本公衆衛生雑誌58(10)pp.03906を参照のこと)。
表は、ある特定の病態を示した急性心筋梗塞患者に対する、入院後7日以内の心臓カテーテル検査(CATH)実施の有無別に、患者属性を示している。この表によると、RCTではその性格上当然ではあるが、CATHの実施有無はそれぞれ462人、458人とほぼ同数となっている。他方、観察研究であるCCPでは、全調査対象者1万7473人のうち5866人(約34%)のみが当該期間内にCATHを受けている。次に、代表的ないくつかの患者属性についてみると、RCTでは年齢と性別ともに、実施の有無による統計的に有意な差は観察されず、バランスしている。他方、CCPでは年齢が若く、男性の方がCATHを受ける確率が高い傾向にあることがわかる。
次に、RCTでは実施の有無によるインスリン糖尿病、心筋梗塞の既往、神経障害等、リスク要因や副疾患の有無、そして

表 RCT対観察研究(CCP):入院後7日以内心臓カテーテル検査実施の有無別の患者属性

変数の定義	RCT		観察研究	
	実施有 (N=462)	実施無 (N=458)	実施有 (N=5,866)	実施無 (N=11,607)
年齢(標準偏差)	62.0 (10.0)	61.0 (10.0)	71.4 (7.5)	77.4 (9.1)
女性(%)	3.0	2.2	38.2	51.3
高血圧	56.7	51.5	64.3	66.9
インスリン糖尿病	24.9	27.3	28.3	34.5
心筋梗塞の既往	43.1	43.0	5.1	7.3
神経障害	11.5	11.8	1.1	7.1
ペーシング頻出率(EF)(標準偏差)	53.0 (15.0)	50.0 (14.0)	48.3 (12.8)	44.2 (13.3)

出所: RCTについては、Boden WE et al. (1998) "Outcomes in Patients with Acute Non-Q-Wave Myocardial Infarction Randomly Assigned to an Invasive as Compared with a Conservative Management Strategy". New England Journal of Medicine. 338(25):pp.1785-1792.

急性心筋梗塞の重症度を示す指標の一つであるペーシング駆出率について、統計的な有意差はほとんど認められないのに対し、CCPでは状態のよい患者の方がCATHを受ける比率が

高いことが示されている。以上の結果から、観察研究では入院後7日以内のCATHの実施に象徴されるような積極的治療の実施に、明らかに偏り(選択バイアス)があることがわかる。

仮に、こうしたデータに基づき、日本でいえばNDBに代表されるような観察研究を用いた分析を行ったとして、CATHの実施有の方が実施無に比べ、有意に良好なアウトカムが得られたとしよう。しかし、患者属性で見ると、CATHの実施有にリスクの高い重篤な患者が多い傾向にあることから、実施有と実施無とのアウトカムの差が、純粋に積極的治療の効果なのか、それとも、CATHの実施有の患者属性の良好な健康状態が影響しているのかを厳密に識別しない限りは確たるエビデンスとは言えない。つまり、もともと状態が悪く、増悪しないしは死亡する確率が高い患者に對しては、積極的治療を行わないアウトカムと積極的治療との間に逆相関が存在することになる

観察研究をどう調整するか?

前段でのべた観察研究による結果の偏りを調整するためには、観察研究によって得られたサンプルを疑似的にランダムマイズするための統計的な工夫を施す必要がある。代表的なやり方として、たとえば、診療報酬の改定等の制度変更があつた場合、それを「自然実験」としてとらえ、その前後で患者や医療従事者の行動がどのように変容するかを観察するという手段が考えられる。いずれにしても、こうした手法を医療情報へ応用

するためには、患者自身も含め、医療情報に携わるすべて組織や関係者との協力的体制を築きながら、医療情報の全国規模でのデータベース化、及びそれを活用するための法的整備を行わなければならない。
第一に、専門家による学際的な研究グループを組織し、医療

評価に対するコンセンサスに基づく医療情報の標準化を促すため、社会全体の理解と協力が得られるよう働きかけること、また、標準化された情報を常時見直すことができる物理的システムと人的協力体制をつくりあげること、そして、そうした取組みを、国家レベルで総合的かつ持続的に行うことである。

第二に、ある特定の医学研究あるいは政策課題に対する問題設定を行うにあたって、ステイ・デザインを企画する段階から、実際の分析結果のまとめを行う最終的な作業に至るまで、すべての段階におけるクオリティ・チェックと、政策の結果と政策過程自体を常時見直す手続きを怠らないことである。

第三に、患者個人を含めデータの提供者と被提供者の双方が共通の利益に向け、長期的な信頼関係を築くことができるような環境を整備するため、情報の運用と管理システムにおける透明性のあるルールを構築し、そのための人材育成を促進することが急務であろう。