

# 保健・医療分野における研究の評価基準

## 定量的基準と定性的基準の再構築

ミヤタ ヒロアキ カイ イチロウ  
宮田 裕章\* 甲斐 一郎<sup>2\*</sup>

保健・医療分野における定量的研究と定性的研究のパラダイム間のつながりについての議論は多くの場合混乱したものであり、散乱する用語と議論によって、概念が不明瞭かつ認識困難となってしまう。このような状況で、厳密さを確立するための評価基準を再構築することは重要である。本研究では、定量的研究・定性的研究における評価基準の背景にある認識論を考察し、多重選択のパラダイムにより、状況や目的の設定によって、どのような評価基準を選択する必要があるのかを議論した。定量的研究・定性的研究の比較議論の中で最も重要な概念はLincolnらが指摘した妥当性/信用性、信頼性/一貫性、客観性/確証性、一般化可能性/転用可能性の対比である。これらの対比点はそれぞれ、観察の枠組みの設定、観察における安定性の仮定、観察者と被観察事象の影響、知見の適用範囲についての認識の違いであると考えられる。ただし現状でも、全ての定性的研究の手法が後者のパラダイムを選択している訳ではなく、一部の安定性など前者の仮定が部分的に成り立つ範囲において、そのパラダイムによる評価を行うことは有用であると考えられる。定量的研究においても、全ての研究で解釈の余地がない枠組みや普遍的な一般化、観察の全プロセスへの安定性、などの前提が仮定できる訳ではない。適用範囲外への知見の外挿や、安定性が確保できない範囲について、後者のパラダイムを用いた評価を行うことは厳密さを高める上で有用である。定量的・定性的の研究手法に関わらず研究者は各々の状況について、1. 観察の枠組み、2. 観察における安定性、3. 観察者と被観察事象の影響、4. 知見の適用範囲、の各項目に関してどのような前提を設定することができるか認識し、設定に応じたバランスによって厳密さの評価を行う必要があると考えられる。

**Key words** : 妥当性, 信頼性, 一般化可能性, 定量的研究, 定性的研究, 評価

### 1 定量的研究と定性的研究の関係と課題

定量的研究では、多くの場合データを集める際に連続/不連続なデータとして統計的分析にかけることができるようなカテゴリーをあらかじめ想定し、概念化を行う。この場合、尺度の正確さは数えられるものに注目することによって得られる。一方で定性的データは、状況・事柄・人々・相互関係・目にみえる行動に関する詳細なデータから構成され、多くの場合は制限のない語りの形で収集される。定性的研究では定量的研究におけるテストの選択肢のように、あらかじめ想定され

たカテゴリーや標準化されたカテゴリーを用いることは少ない<sup>1)</sup>。

定性的研究は、保健・医療分野の研究で、近年広く用いられるようになってきているが、もともとは社会科学の分野で使われていた手法であるために、生物学や自然科学を基礎にしている保健・医療専門職や研究職にはとっつきにくい面がある<sup>2)</sup>。科学者・科学史家のKuhn<sup>3)</sup>の「定量的な予測は定性的な予測より好ましい」という見解や、Bernstein<sup>4)</sup>らが定性的データと定量的データの両方を集めた調査を、方法論の価値において定量的調査のみを集めた調査よりも低いランクにおいたこと、なども定性的研究の受け入れを妨げている要因であろう。

社会科学の学問分野が、主として定量的方法を用いて社会構造を明らかにすることを重視する立

\* 早稲田大学人間科学学術院

<sup>2\*</sup> 東京大学大学院医学系研究科健康科学・看護学専攻  
連絡先: 〒155-0033 東京都世田谷区代田 6-7-2  
早稲田大学人間科学学術院 宮田裕章

場と、主として定性的手法を用いて社会活動の様子やその意味を理解することを重視する立場に分かれているために、定性的研究と定量的研究が互いに対立する関係にあるかのように強調されることが多い<sup>5,6)</sup>。しかしながら今では、「定性的研究と、定量的研究を区別することは何の役にも立たず、区別自体がいい加減なものだ」という見解は、社会学<sup>7,8)</sup>のみならず保健・医療サービスの専門分野<sup>9)</sup>でも浸透しつつある。

アンケート調査に使う用語や言葉を十分に幅広く抽出するために定性的調査を行うなど、定量的研究をする前に、その先行的な研究として定性的研究が有用となることもある。また定量的研究のデータを詳しく解釈し理解するのに、定性的研究が役立つこともある。これは定量的調査によって把握された調査項目に大きなばらつきが生じていた時に、その理由を定性的アプローチで明らかにする、というような研究手順を例として挙げることができる。このように、定量的研究と定性的研究を、お互いに対立するものとしてではなく、補完しあうものとして利用することは重要な研究視点である。加えて、定性的研究は定量的研究を補うだけでなく、定量的方法では分析が難しい問題に対して単独で用いることもできる。

上記のように定性的方法と定量的方法を多重的に利用する場合には、それぞれのデータを公平に評価する必要がある<sup>1)</sup>。一方で、定量的研究と定性的研究を個別の研究として報告する場合であっても、体系的なレビューなどで問題を概観する場合には、両者の研究成果を複合的に用いる必要がある<sup>10,11)</sup>。したがって定量的研究と定性的研究における評価基準を再検討することは重要な課題といえる。

## II 研究の評価基準に関する議論の経過

定量的研究における厳密さの評価基準として広く用いられているのは妥当性 Validity, 信頼性 Reliability, 一般化可能性 Generalizability という概念である。しかしながら近年、客観的な現実や確証的な知覚を前提としない定性的研究の立場からは、これらの基準についての絶対性が疑問視されるようになってきている<sup>12)</sup>。また、妥当性、信頼性、一般化可能性という用語を定性的研究のパラダイムにおいて使うべきかどうかという議論は20

年以上に渡っており<sup>13~18)</sup>、基準の確立を標榜するガイドラインについて、領域をこえた活発な検討が行われている<sup>19)</sup>。ただし定量的研究と定性的研究のパラダイム間のつながりについての議論は多くの場合混乱したものであり、散乱する用語と議論によって、概念が不明瞭かつ認識困難となってしまう<sup>20)</sup>。

定性的研究の方法論に大きな変革が訪れたのは1980年代であり、この時、研究のための新しい概念が誕生した。妥当性、信頼性、一般化可能性にかわるものとして Lincoln らは、定量的研究のパラダイムの境界の外側から定性的研究のために信憑性 trustworthiness という基準を提案したのである<sup>13)</sup>。Lincoln ら当人も認識していたように彼らの基準は不完全なものであったが、それまで絶対的であった伝統的な定量的研究のパラダイムへの対抗馬を立てたことは大きな意義があった。Lincoln らが提案した信用性 Credibility, 一貫性 Dependability という概念は革新的かつ挑戦的であり、研究の厳密さについての現状の議論の礎を築いたといつてよい<sup>21)</sup>。その後 Lincoln らは信憑性の基準を再検討し信用性、一貫性に、転用可能性 Transferability, 確証性 Confirmability の2つを加えた4つの概念を提案した<sup>22)</sup>。

定性的研究をめぐる基準について主流を占める議論は、定性的研究には新しいアプローチが必要とされていて、定量的研究のアプローチに結びついた基準は棄却せねばならないというものであり、これがしばしば妥当性・信頼性という基準を棄却することに結びついている<sup>23,24)</sup>。また他に定量的研究の基準を棄却する理由としては、妥当性・信頼性という用語が定量的研究のパラダイムで必ず登場する一方で、定性的研究のパラダイムではそれほど登場しないということが挙げられている<sup>25,26)</sup>。しかし一方で、Morse<sup>20)</sup>はこのように定量的研究の基準を徹底的に排除することは、定性的研究が科学として棄却されることにつながると主張し、現在の流れに警鐘を鳴らしている。Morse は科学において厳密さは重要な要素であり、その基準である妥当性、信頼性を棄却することは、定性的研究が厳密さを放棄するものとみなされると考えている。

このような状況で、定性的研究と定量的研究の評価基準を再検討し、両者に共通する枠組みを検

討することは重要である。本研究は、定量的研究における評価基準である妥当性、信頼性、一般化可能性、および定性的研究における評価基準である信用性、一貫性、転用可能性、確証性についてそれらの背景にある認識論を検討することにより、基準についてその統合の可能性を考察する。

### III 多重選択のパラダイムによる視点

本研究では、Patton<sup>27)</sup>が提示した有用性に基づいた多重選択のパラダイムにその視点に置いて検討を行う。多重選択のパラダイムは、各状況や目的に対して適切な、さまざまな方法があることを認めるものである。したがってこの視点は、有用性を根拠に特定の認識論の真实性を主張するという、従来の実用主義とは異なるものである。

本研究では、定量的研究・定性的研究における評価基準の背景にある認識論を考察し、多重選択のパラダイムにより、状況や目的の設定によって、どのような評価基準を選択する必要があるのかを議論する。このように全ての科学的方法論の強みと限界を認識した上で多様な視点からの批評的で網羅的な現象研究を行うことは、批評的多元主義の立場からも奨励されている<sup>28)</sup>。しかしながら、多くの研究者が有用性によって視点を選択するようになった状況でも、妥当性・信頼性の基準は依然として不明瞭である<sup>29)</sup>。また方法論がはっきりしないことは厳密さの欠如につながる。本研究ではLincolnらが対比した、妥当性と信用性、信頼性と一貫性、客観性と確証性、一般化可能性と転用可能性の基準を照らし合わせて、パラダイムの検討を行う。

### IV 妥当性/信用性

Cronbach<sup>30)</sup>は、妥当性について「特定の方法から生まれるデータの解釈が、測ろうとしている内容をどれくらい適切に測っているかという基準」と説明している。また定量的研究における妥当性の概念の他の説明としては、「研究枠組みが測定しようと意図したものを本当に測定しているかどうか、研究結果がどの程度真実であるか」というものがある<sup>31)</sup>。妥当性には複数の類型があり基準関連妥当性、構成概念妥当性、内容妥当性などが例として挙げられる<sup>32~35)</sup>。基準関連妥当性とは、測定値を客観的に評価するために、外的な

基準を設けて測定値と照らし合わせることによって評価されるものである。構成概念妥当性は、ある理論的概念を表すように計画された多重指標間の関係と理論的に関連する外的変数群が、方向・強度・一致性の点で似ているかどうかを評価するものである。内容的妥当性は経験的な測定が内容の特殊な領域をどの程度反映しているかを評価するものである。しかしながら誤った内容の枠組みで観察するという過誤はしばしば重大な問題で、しばしば取り除くことが難しい問題である<sup>36)</sup>。Bohrnstedt<sup>37)</sup>は、内容的妥当性の概念をそれを評価する厳密な手段がないという理由で拒否している。このように妥当性は設定された観察枠組みについての評価パラダイムであると考えられる。また研究を開始するにあたり観察枠組みを予め設定することが可能であるという認識を前提としていることが示唆される。

これに対してLincolnらは、定性的研究では事実の一つではなく、人間の精神によって構成される複数のものであると設定している<sup>22)</sup>。この場合、研究者による事実の再構成が当事者にとっても信用にたるものであるという基準によって厳密さを確保する。信用性は研究者が表現する対象者の視点と対象者本人の視点との適合性についての概念であり<sup>38)</sup>、説明が記述に適合するかどうか、記述が確かなものかどうかについての疑問を提起するものである<sup>39)</sup>。信用性は情報源・方法・調査者について複数を組み合わせるトライアングレーションという手法や、場になれず常に注意深く観察すること、ネガティブケースの分析などによってその質が確保される<sup>17)</sup>。このことから信用性は、観察によって新たに形成される枠組みについての評価パラダイムであると考えられる。また研究を開始するにあたり観察枠組みを予め設定することが難しいという認識を前提としていることが示唆される。

妥当性/信用性のパラダイムに関しては、観察の枠組みをどのように設定するかについての認識に大きな違いがあると考えられる(表1)。枠組みを予め設定することが可能であるという認識がある場合には妥当性というパラダイムが評価に用いられている。この場合、枠組みを設定するということは固定された1側面から記述するという意味で、仮説を検証するというよりも広義な捉え方

表1 評価パラダイムとその背景にある認識論の関係

属性	属性に対する認識論	評価パラダイム	活用方法
観察の枠組み	研究開始前に設定することが可能な枠組み → 予め設定した枠組みを用いて観察を行う	妥当性	観察において、枠組みを予め設定した領域については妥当性の基準で、枠組みを新たに生成する領域については信用性で評価を行う
	研究開始前に枠組みを設定することが難しい → 観察によって新たな枠組みを生成する	信用性	
観察における安定性	事象や方法に安定性を仮定することが可能 → 結果についての評価によって安定性を検討する	信頼性	観察において、安定性が仮定できる領域については、信頼性の基準で評価を行い、安定性の確保が難しい点については一貫性の基準を適用する
	事象や方法に安定性を仮定することが困難 → 研究の手順についての安定性の確保を試みる	一貫性	
観察による影響	観察や介入に中立性を仮定することが可能 → 手順の執行についてのチェックが行われる	客観性	観察者と被観察事象の間に中立性が仮定できる領域について客観性の基準を用いて確保を試み、難しい領域については、確証性の基準によって厳密さの評価を行う
	観察や介入に中立性を仮定することが難しい → 活動の影響過程や帰結を考慮して評価を行う	確証性	
知見の適用範囲の設定	知見を直接適用することが可能である領域に対して → 一般化に向けた検討を行う	一般化可能性	知見を一般化する範囲を設定し、その範囲への知見の適用を一般化可能性の基準で評価する。範囲外の異なる集団に対しては転用可能性の基準を用いて評価を行う
	知見を直接適用することが難しい領域に対して → 異なる集団に対する外挿に情報が必要とされる	転用可能性	

である。また、あらかじめ枠組みを設定することが難しいという認識がある場合には信用性というパラダイムが評価に用いられている。また、観察において定量的研究が解釈の幅が狭い数字を用い、逆に定性的研究は解釈に幅がある言語を用いていることから、多くの場合定量的研究が妥当性、定性的研究が信用性を用いた評価が主体となると考えられる。

ただ定性的研究においても対象地域の中に入って生活する参与観察を用いた研究では仮説検証を主眼にすることも多く<sup>40)</sup>、この場合予め設定した枠組みが存在するため、妥当性という基準においても評価を行うことは必要であると考えられる。加えて仮説生成を主眼にした定性的研究であったとしても、データ収集におけるインタビューの質問内容や構成などある程度の枠組みを設定することもある。このような場合、事実の再構成について信用性という基準で評価を行うのと同様に、設定した枠組みに対して妥当性という基準で評価を行うことは有用である。

定量的研究においても同様で、全ての研究にお

いて厳密な枠組みを予め設定することが可能なわけではない。とくに探索的因子分析やデータマイニングの手法は、観察の枠組みに幅があり信用性という基準による厳密さの確保は有用である可能性がある。しかしながら現状の信用性の手順は、言語を用いた分析に限定されたものが多く、数量データに関する信用性の手順の形成は今後の課題となるであろう。加えて、定量的研究における質問紙調査では、言語を用いて概念を形成する行程が必ず存在する。この場合は研究者による調査票の作成は事実の再構成の過程であり、信用性の基準を適用する必要性が生じる。先に挙げたように定量的研究における内容的妥当性は重大な問題であり、現状の評価手順だけでは厳密さを確保することが難しく、信用性パラダイムを用いた評価は有益であると考えられる。定量的研究における信用性のパラダイムの活用には今後更なる検討が望まれる。

妥当性/信用性は観察における枠組みの設定に関わるパラダイムである。枠組みを形成するプロセスについて評価する場合には信用性の基準を、

設定した枠組み自体を評価する場合には妥当性の基準を用いる必要がある。保健・医療の研究において、研究個別の枠組みの設定に応じて妥当性/信用性のパラダイムを並行的に用いることは厳密さを高める上で有用であると考えられる。

## V 信頼性/一貫性

信頼性とは、実験や検査など任意の測定手段でデータ収集を反復した場合、どの程度同じ結果を示すかという事柄に関係している<sup>34)</sup>。一般に、どんな現象の測定でも、ある量の偶然誤差が含まれ、誤差がまったくない測定は現実的には達成不可能であると考えられている。Kirkら<sup>36)</sup>は、更に定量的研究における3つのタイプの信頼性を同定した。ドンキホーテ的 quixotic 信頼性、共時的 synchronic 信頼性、通時的 diachronic 信頼性という三つの基準を提出している<sup>22)</sup>。ドンキホーテ的信頼性とは、ある方法によって全く同じデータがいつでも得られるか、共時的信頼性とは、異なったデータ収集のツールを用いた場合に測定や観察の結果が一貫しているか、通時的信頼性とは、ある現象の測定結果や観察が時間経過の中で安定しているか、を指している。Kirkらは、ドンキホーテ的信頼性の基準を有用でないと棄却している一方で、定性的研究においては、共時的・通時的信頼性の基準が満たされないケースも存在することを指摘している。信頼性を検討する方法としては、再テスト法（同じ個人に対して、一定期間をおいて同じテストを実施）、折半法（全項目が半分ずつに分けられ各々の得点の相関を求める）、内的整合性（項目間の相関係数の平均をとる）が挙げられる<sup>34,41)</sup>。このように信頼性のパラダイムは観察結果について評価を行うものであり、観察事象や観察方法について安定性を仮定することが可能であるという認識に基づいていると考えられる。

観察する事象が、時間や環境、観察手法によって変化する可能性が高い場合には、安定性を仮定することは難しくなる。この場合信頼性ではなく、一貫性という概念の下、研究を通じたデータ収集とその分析過程を一貫させることによって、結果と解釈の根拠を明確にすることが必要となる<sup>22)</sup>。また、研究者は調査の過程が論理的かつ追跡可能で明瞭に文書化されていることを保証する

必要がある<sup>38)</sup>。一貫性は主として監査 audit trail によって確認される。ここでいう監査とは調査する人によって記録された調査の経過を系統的に第三者が点検する手続きをいう<sup>42)</sup>。たとえばインタビュー調査で点検対象となるのは、録音テープや、トランスクリプト、インタビューガイド、データ分析時に用いたカテゴリーや仮説のリストなどである。ここでは一貫性のパラダイムは観察方法やデータ分析など研究過程の一部分についての評価を行うものであり、先に示したように観察方法や事象に安定性することが難しいという認識を前提にしている。

信頼性/一貫性のパラダイムに関しては、観察における安定性をどのように仮定するかについての認識に大きな違いがあると考えられる（表1）。観察対象とする事象や観察方法について安定性を仮定することが可能であるという認識がある場合に、信頼性というパラダイムが用いられるのに対して、観察する事象や観察方法に安定性を仮定することが難しいという認識がある場合には、一貫性というパラダイムが評価に用いられる。実験室内など変化する指標が限定され、比較的條件を統制することが可能な場合の事象の観察は信頼性のパラダイムに親和性があり、現実の社会環境における人間の感情や社会的な出来事など、変化する指標を限定することが難しい場合の事象の観察では一貫性のパラダイムに親和性がある可能性が高い。

観察の枠組みについて先に挙げた例と同様に、観察における安定性についても、定量的研究で常に完全に仮定することが可能で、定性的研究で常に全く仮定できないとは限らない。たとえば、精神障害は症状に個人差があり病態に安定性を仮定することが一般に難しい。このような精神障害を扱う分野ではDSM<sup>43)</sup>などの診断基準に基づいて障害を定義することにより、各個人の症状と診断名の間に一貫性の確保を試みている。このように診断を確定した後に、投薬の効果を定量的に検討する、症状の特徴を定性的に検討するなどの研究が行われている。意識的にこのパラダイムに基づいたかどうかは不明であるが、一貫性のパラダイムを活用した一つの例であると考えられる。このように保健・医療の研究においては、単純に安定性を仮定できるか出来ないかという認識

によって単一のパラダイムを選択するのではなく、観察手法や観察過程、被観察事象、データの分析などの各要素について、どのような点に安定性を仮定でき、どの部分に仮定することができないかという判断を行う必要がある。つまり安定性が仮定できる部分については信頼性のパラダイムで評価を行い、安定性を仮定することが難しい部分については一貫性のパラダイムで評価を行うことが研究の厳密さを確保する上で有用である可能性がある。

## VI 客観性/確証性

観察者が被観察事象に対して取るべき距離についても主として二つの立場がある。一方は接近しすぎると客観性が損なわれるとの立場であり、もう一方は離れすぎると観察や理解が足りなくなるという立場である<sup>1)</sup>。定量的研究は多くの場合前者の立場をとり、客観性を保つために被観察事象と距離を置く。この客観性の前提としては、①研究のデータと現実が同質性をもつと考えられる、②観察者と被観察者に距離を持つことは可能、③研究者の価値は研究に反映されないという条件がある<sup>2)</sup>。このように中立性が仮定できる場合には客観性の基準の下に、手続きにおける人為的ミスの及ぶ可能性の検討や二重盲検法、三重盲検法などを行うことによって厳密さを確保する。

一方で定性的研究はしばしば、直接的な経験からくる感情移入や共感なくしては、十分に理解できない問題を取り扱う。この目的を達成する上で、たとえば、研究の対象者に接近してものごとの細部まで注意を払い、彼らと長い時間を一緒に過ごし親密な問柄となって信頼を得る事が必要となることもある<sup>4)</sup>。当然ながら、この時研究者のデータ収集が、被観察者に何らかの影響を与えると考えられる上に、研究者自身も被観察事象に影響をうける可能性があり、客観性を前提とすることは難しい。したがってこの場合重要となるのは、研究の結果がデータに基づいているかを確認することである。定性的研究では確証性の基準のもとに、データや知見の解釈が研究者の想像による虚構ではなく、明瞭にデータから導き出されたものであることを示すことが推奨されている<sup>2)</sup>。確証性を確認する方法としては、データの知見や解釈が研究者の想像ではなく、明瞭にデータから

導き出されたものであることを示すことや、研究方法について詳細に明記し、どのようにデータが集められ・要約され・最終的な結論が導かれたのか辿ることができるようにすること等が挙げられる<sup>29)</sup>。

以上から客観性/確証性のパラダイムに関しては、観察者と被観察事象の間に生じる影響についての認識に大きな違いがあると考えられる(表1)。観察者と被観察事象の間に中立性を仮定することが可能であるという認識がある場合に客観性というパラダイムが評価に用いられるのに対して、観察者と被観察事象の間に中立性を仮定することが難しいという認識がある場合に確証性というパラダイムが評価に用いられる。測定機器を用いて物体の動きを観察する場合など、人間の関与する領域が比較的限られる研究においては、ほとんどの場合で中立性が前提とされており、この基準による評価自体に感心が低い傾向にある。これは定量的研究の評価基準に客観性がしばしば含まれないことによっても示されている。

人間を対象とする保健・医療の領域においては、定量的研究においても研究が対象者に影響を及ぼすことは少なくはない。面接法によるデータ収集で人間同士の相互作用によりさまざまな影響が発生することは当然であるが、質問紙を用いた場合でも質問事項の表現や設問の構成などが、対象者の回答に影響を及ぼす可能性はある。また研究の成果によって実績を挙げる研究者本人以外にも、被観察事象である回答者や研究費の出資元、共同研究者が研究結果について影響を受ける場合がある。近年、一部の科学ジャーナル誌は研究について本編外の部分で出資元や共同研究者についてのコメントを要求することがあるが、この基準についての意識が広まっているとはいえない。今後はデータ収集やデータの解釈における観察者と被観察事象の影響のみではなく、研究の背景にある出資者や所属組織の影響について検討することも、厳密さを高める上で有用であるかもしれない。

## VII 一般化可能性/転用可能性

Rothman<sup>45)</sup>は、科学における一般化を考える有用な方法として、科学理論を緻密化する作業として一般化を考えることであると述べている。検証で生き残った理論は自然に関する一般的言明と捉

えることができ、直接研究対象となっていない人々や状況では、何が起こると期待できるかを教えてくれるものであるという。この一般化について Cronbach<sup>46)</sup>は社会科学の一般化の概念は自然科学のそれとは異なるものであると指摘し、対象とする現象が多様かつ状況による変化が激しい場合にまずすべきは、異なる状況に遭遇するたび状況を記述し、影響を解釈し、その状況に特異的な説明要因を考慮に入れていくことであると述べている。このように局所的状況を適度に重要視すれば、社会科学における一般化というものは、結論ではなく作業仮説になるということである<sup>46)</sup>。保健・医療においては Rothman が前提とするような知見を一般的な言明として適用可能な状況もあれば、Cronbach が指摘するような知見を限られた範囲における作業仮説に留める必要がある状況も存在する。したがって、一般化を考慮する場合には、研究の知見をどのような範囲に適用可能であるのかを判断することは重要であると考えられる。定量的研究・定性的研究の双方で知見の一般化を判断する基準として Shadish<sup>47)</sup>が提案したのが、①近似的類似性の原則(背景が類似している)、②無関係不均質の原則(異なる条件でも知見が通用する)、③区別的妥当性の法則(研究知見が生み出されるために本質的に必要な構成要素であると示すことができる)、④経験的内挿外挿の法則(経験的に、知見が別の人や設定、治療、時代に適用できるか、全くできないかが判断できる)、⑤説明の原則(ある変数群が別の変数群と、どのように関係して、どのような相互作用が起きているか示せるとき)、である。このようにある範囲に対して知見の一般化が可能であるという認識がある時、一般化可能性のパラダイムに基づいて知見を適用する集団を定義し、その範囲への一般化を評価する。

一方、定性的研究のパラダイムにおいては単純な正解だとか真実の解釈がある訳ではなく、知見の一般化を迫り出すのとは別の認識を持つ必要があるとする見解もある。Donmoyer<sup>48)</sup>は、定性的研究では個々に主観的な意味合いを有するので、伝統的な一般化可能性を捨てる必要があると主張している。これに関して、文脈を送る・受け取る類似の状況において、送り手文脈から出た作業仮説を受け手の文脈でも応用するため転用可能性

という基準が提案されている<sup>22)</sup>。Lincoln らは時と場所が違って同じ知見が得られるかどうかは、文脈を送る/受け取ることに依存する経験的な問題であると述べている、転用可能性の改善には、読み手が潜在的な応用性について判断できるに十分な記述を知見に含めること、知見がさらに有益となるための検証のありかについての提言を行う<sup>29)</sup>という方法が提案されている。このように知見の一般化が難しいという認識がある時、転用可能性のパラダイムに基づいて異なる集団への外挿を評価する。

一般化可能性/転用可能性のパラダイムに関しては得られた知見の適用範囲についての認識に大きな違いがあると考えられる(表1)。ある集団内において知見を一般化することができるという認識がある場合については、一般化可能性というパラダイムを用いて評価を行う一方で、知見の一般化は困難であり異なる集団への外挿が必要であるという認識がある場合には、転用可能性という評価基準を用いる。技術や理論についての基礎研究に比較して、応用分野の研究は社会的制度や地域環境、対象者の個人特性などに影響を受けやすいため一般化可能性を広く確保することが難しい。ただし、当然ながら応用分野の研究は、対象者への知見の直接の還元性という点で有益となる可能性が高く、知見の適用範囲のみで研究の質を判断することは避けるべきである。

介入プログラムを試行しその結果を評価する定量的研究を例に考えると、この研究の知見はプログラムが試行予定である地域における一般化が必要である。一方で、プログラムの前提となる制度や対象者の特性が異なる他国の地域などにおいては、同様のプログラムを単純に適用することは難しい。この場合、一般化可能性の限界に言及するだけでなく、他地域で適用する場合の知見の転用可能性について記述することは、知見の有用性を高めると考えられる。定量的研究における転用可能性の記述については今後手順を確立する必要がある。

定性的研究においても症例研究など個別の事例についての詳細な報告を行う研究では、一般化可能性の基準で研究を評価することは有用ではないかもしれない。しかしながら定性的研究においても、グランデッドセオリーを用いた研究は、知見

の適用範囲として中範囲を標榜するものであり、限定した範囲での一般化可能性を評価する手順の提案を行っている。このように定性的研究でも知見の適用範囲を設定した上で、その限られた中で知見の一般化について言及することは、有意義であると考えられる。

一般化可能性/転用可能性は知見の適用範囲の設定に関わるパラダイムである。研究の手法が定性的・定量的かどうかに関わらず、研究の知見の適用範囲について認識し設定することは重要である。普遍的な一般化が仮定できる、あるいは他のケースへの適用可能性について全く言及できないというような適用範囲が極端となるような設定ではない限り、範囲内の一般化と範囲外への転用可能性について言及することは知見の有用性を高める上で意義があることであると考えられる。

## VIII 結 語

本研究では、定量的研究・定性的研究における評価基準の背景にある認識論を考察し、多重選択のパラダイムにより、状況や目的の設定によって、どのような評価基準を選択する必要があるのかを議論した。定量的研究・定性的研究の比較議論のなかでの最も重要な概念はLincolnらが指摘した妥当性/信用性、信頼性/一貫性、客観性/確証性、一般化可能性/転用可能性の対比である。これらの対比点はそれぞれ、観察の枠組みの設定、観察における安定性の仮定、観察者と被観察事象の影響、知見の適用範囲についての認識の違いであると考えられる。ただし現状でも、全ての定性的研究の手法が後者のパラダイムを選択している訳ではなく、一部の安定性など前者の仮定が部分的に成り立つ範囲において、そのパラダイムによる評価を行うことは有用であると考えられる。定量的研究においても、全ての研究で解釈の余地がない枠組みや普遍的な一般化、観察の全プロセスへの安定性、などの前提が仮定できる訳ではない。適用範囲外への知見の外挿や、安定性が確保できない範囲について、後者のパラダイムを用いた評価を行うことは厳密さを高める上で有用である。定量的・定性的の研究手法に関わらず研究者は各々の状況について、1. 観察の枠組み、2. 観察における安定性、3. 観察者と被観察事象の影響、4. 知見の適用範囲、の各項目に関してど

の様な前提を設定することができるか認識し、設定に応じたバランスによって厳密さの評価を行う必要があると考えられる。

(受付 2005. 8. 8)  
(採用 2006. 3.31)

## 文 献

- 1) Patton MQ. Utilization-Focused Evaluation: The New Century Text. Thousand Oaks, CA: Sage Publication, 1997.
- 2) Pope C, Mays N. Qualitative methods in health research. London: BMJ Publishing Group, 2001.
- 3) Kuhn T. The Structure of Scientific Revolutions. Chicago: University of Chicago Press, 1970.
- 4) Bernstein I, Freeman HE. Academic and Entrepreneurial Research: Consequences of Diversity in Federal Evaluation Studies. New York: Russell Sage, 1975.
- 5) Mechanic D. Medical sociology: some tensions among theory, method and substance. Journal of Health and Social Behavior 1989; 30: 147-60.
- 6) Pearlin L. Structure and meaning in medical sociology. Journal of Health and Social Behavior 1992; 33: 1-9.
- 7) Abell P. Methodological achievements in sociology over the past few decades with special reference to the interplay of qualitative and quantitative methods. In: Bryant C, Becker H, eds. What Has Sociology Achieved. London: Macmillan, 1990.
- 8) Hammersley M. What's Wrong with Ethnography: Methodological Explorations. London and New York: Routledge, 1992.
- 9) Barbour RS. The case for combining qualitative and quantitative approaches in health services research. Journal of Health Services Research and Policy 1999; 4: 39-43.
- 10) Bravata DM, McDonald KM, Shojania KG, et al. Challenges in systematic reviews: Synthesis of topics related to the delivery, organization, and financing of health care. Annals of Internal Medicine 2005; 142: 1056-1065.
- 11) Mulrow C, Langhorne P, Grimshaw J. Integrating heterogeneous pieces of evidence in systematic reviews. Annals of Internal Medicine 1997; 127: 989-995.
- 12) Klave S. The social construct of validity. Qualitative Inquiry 1995; 1: 19-40.
- 13) Guba EG, Lincoln YS. Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches. San Francisco, CA: Jossey-Bass, 1981.
- 14) Sandelowski M. The problem of rigor in qualitative



- research. *Advances in Nursing Science* 1986; 8: 27-37.
- 15) Mishler EG. Validation in inquiry-guided research. *Harvard Education Review* 1990; 60: 414-422.
  - 16) Lather P. The validity of angles: Interpretive and textual strategies in researching the lives of women with HIV/AIDS. *Qualitative Inquiry* 1995; 1: 41-68.
  - 17) Lincoln YS. Emerging criteria for qualitative and interpretive research. *Qualitative Inquiry* 1995; 3: 275-289.
  - 18) Morse JM, Swanson JM, Kuzel AJ. *The Nature of Qualitative Evidence*. Thousand Oaks, CA: Sage, 2001.
  - 19) Mays N, Pope CP. *Quality in Qualitative Health Research*. London: BMJ Books, 2001.
  - 20) Morse JM, Swanson JM, Barrett M, et al. Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods* 2002; 1: Article2.
  - 21) Tobin GA. Methodological rigour within a qualitative framework. *Journal of Advanced Nursing* 2004; 48: 388-396.
  - 22) Lincoln YS, Guba EG. *Naturalistic Inquiry*. Thousand Oaks, CA.: Sage, 1985.
  - 23) Pech E, Secker J. Quality criteria for qualitative research: Does context make a difference? . *Qualitative Health Research* 1999; 9: 552-558.
  - 24) Whittemore R, Chase SK, Mandle CL. Validity in qualitative research. *Qualitative Health Research* 2001; 4: 522-537.
  - 25) Altheide D, Johnson J. Criteria for Assessing Interpretive Validity in Qualitative Research. In : Denzin N, Lincoln Y, eds. *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage, 1994.
  - 26) Leninger M. Evaluation Criteria and Critique of Qualitative Research Studies. In: Morse JM, eds. *Critical Issues in Qualitative Research Methods*, Thousand Oaks, Sage, 1994.
  - 27) Patton MQ. *Qualitative Evaluation and Research Methods*. Newbury Park, CA: Sage; 1990.
  - 28) Letourneau N, Allen M. Post-positivistic critical multiplism: A beginning dialogue. *Journal of Advanced Nursing* 1999; 30: 623-630.
  - 29) Miles MB, Huberman AM. *Qualitative Data Analysis*. Thousand Oaks, CA: Sage; 1994.
  - 30) Cronbach LJ. Test Validation. In: Thorndike RL, eds. *Educational Measurement*. Washington, DC: American Council on Education; 1971.
  - 31) Joppe M. The Research Process. Retrieved May 24th, 2006, from: <http://www.ryerson.ca/~mjoppe/rp.htm>.
  - 32) Wainer H, Braun H. *Test Validity*. Hillsdale, NJ: Erlbaum Associates, 1988
  - 33) Carmines EG, Zeller RA. *Reliability And Validity Assessment*. Beverly Hills: Sage Publications, 1979.
  - 34) Nunnally JC. *Psychometric Theory*. New York: McGraw-Hill, 1978.
  - 35) Cronbach LJ. Construct validity in psychological tests. *Psychological Bulletin* 1955; 52: 281-302.
  - 36) Kirk JM, Miller ML. *Reliability and Validity in Qualitative Research*. Beverly Hills: Sage, 1986.
  - 37) Bohrnstedt GW. *Measurement*. New York: Academic Press, 1985.
  - 38) Schwandt TA. *Dictionary of Qualitative Inquiry*. Thousand Oaks, CA: Sage, 2001.
  - 39) Janesick VJ. *The Choreography of Qualitative Research Design: Minuets, Improvisations and Crystallization*. Thousand Oaks, Calif: Sage, 2000.
  - 40) Stake RE. The case study method in social inquiry. *Educational Researcher* 1978; 7: 5-8.
  - 41) Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-334.
  - 42) Schwandt TA. *Qualitative Inquiry: A Dictionary of Terms*. Thousand Oaks, CA: Sage, 1997.
  - 43) American psychiatric association. *Diagnostic and statistical manual of mental disorders*. Washington DC; American psychiatric association, 1994
  - 44) Lofland J. *Analyzing Social Settings*. Belmont, CA: Wadsworth, 1971.
  - 45) Rothman KJ. *Epidemiology: An Introduction*: Oxford University press, 2002.
  - 46) Cronbach LJ. Beyond the two disciplines of scientific psychology. *American Psychologist* 1975; 30: 116-127.
  - 47) Shadish WR. Philosophy of science and the quantitative-qualitative debates: Thirteen common errors. *Evaluation & Program Planning* 1995; 18: 63-75.
  - 48) Donmoyer R. *Generalizability and The Single Case Study*. New York: Teachers College Press, 1990.
-

RECONSIDERING EVALUATION CRITERIA REGARDING  
HEALTH CARE RESEARCH:  
TOWARD AN INTEGRATIVE FRAMEWORK OF QUANTITATIVE  
AND QUALITATIVE CRITERIA

Hiroaki MIYATA\* and Ichiro KAI<sup>2\*</sup>

**Key words** : Scientific criteria, Rigor, Validity, Reliability, Generalizability, Evaluation

Debate about the relationship between quantitative and qualitative paradigms is often muddled and confused and the clutter of terms and arguments has resulted in the concepts becoming obscure and unrecognizable. It is therefore very important to reconsider evaluation criteria regarding rigor in social science. As Lincoln & Guba have already compared quantitative paradigms (validity, reliability, neutrality, generalizability) with qualitative paradigms (credibility, dependability, confirmability, transferability), we have discuss use of evaluation criteria based on pragmatic perspective. Validity/Credibility is the paradigm concerned to observational framework, while Reliability/Dependability refer to the range of stability in observations, Neutrality/Confirmability reflect influences between observers and subjects, Generalizability/Transferability have epistemological difference in the way findings are applied. Qualitative studies, however, does not always chose the qualitative paradigms. If we assume the stability to some extent, it is better to use the quantitative paradigm (reliability). Moreover as a quantitative study can not always guarantee a perfect observational framework, with stability in all phases of observations, it is useful to use qualitative paradigms to enhance the rigor in the study.

---

\* Department of Health and Welfare Policy, School of Health and Welfare Science, Graduate School of Human Science, Waseda University, Japan

<sup>2\*</sup> Department of Social Gerontology, School of Health Sciences and Nursing, Graduate School of Medicine, University of Tokyo, Japan