

《 総 説 》

リハビリテーション患者データベースの二次分析
—プロセス、可能性と限界—

近藤 克則*

Secondary Analysis of the Rehabilitation Patient Database :
Process, Potentials and Limitations

Katsunori KONDO*

Abstract : The Japanese Association of Rehabilitation Medicine (JARM) is developing a Rehabilitation Patient Database (DB). The accumulated number of registered patients exceeded 10000 by March, 2011. The purposes of this article are to describe the process and procedures of secondary analysis and to consider potentials and limitations of the DB to promote the research activities of JARM members. JARM Members who submitted patient data or cooperate with JARM in the secondary analysis are regarded as eligible to use the combined data submitted by many hospitals. A suitable patient dataset should be derived from the DB including stroke, hip fracture, and spinal cord injury, and also patient data from the acute to recovery phase of rehabilitation. Additionally, before paper drafts can be submitted, a reviewing process is needed. The DB holds much potential, because the sample size is large and data were submitted from many hospitals. Since there are inherent limitations in all observational research, many issues such as endogeneities and confounders should be considered carefully to ensure high quality evidence is obtained with validity and reliability using the DB. (*Jpn J Rehabil Med* 2012 ; 49 : 142-148)

Key words : 大規模データベース (large size database), 患者登録 (patients' registry), 二次分析 (secondary analysis), 多施設共同研究 (multi-institutional joint research)

はじめに

日本リハビリテーション医学会がデータマネジメントワーキンググループ (2008) を経て特別委員会を立ち上げ, リハビリテーション (以下, リハ) 患者データベースを中核とするデータ・マネジメント・システムの開発と運用に取り組み始めたのは2010年である。その後, 2011年3月までに, 1万件を超えるデータが蓄積された。

データ・マネジメント・システムは, 図1に示したような流れによって運用されている¹⁾。まず, リハ患者データベースを開発し, 参加病院に患者データを入力してもらい, 個人情報削除後に提出していただき, その登録データを結合する。ここまでが大規模データベースを構築するプロセスである。その後, 蓄積されたデータを活用する局面となる。1つの方法は集計結果をデータ提供病院にフィードバックするレポートの作成であり, もう1つが, 登録データを活用したエビデンスづ

2012年3月5日受稿

* 日本リハビリテーション医学会データマネジメント特別委員会 委員長
日本福祉大学社会福祉学部/〒460-0012 愛知県名古屋市中区千代田5-22-35
Faculty of Social Welfare, Nihon Fukushi University
E-mail : kkondo@n-fukushi.ac.jp

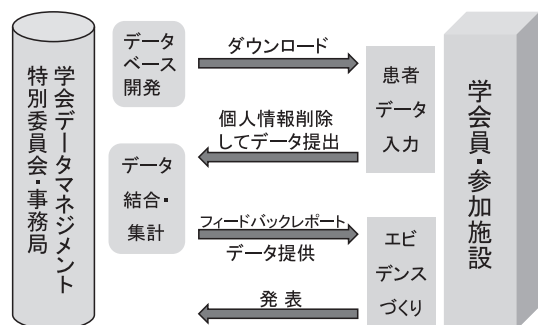


図1 データ・マネジメント・システム

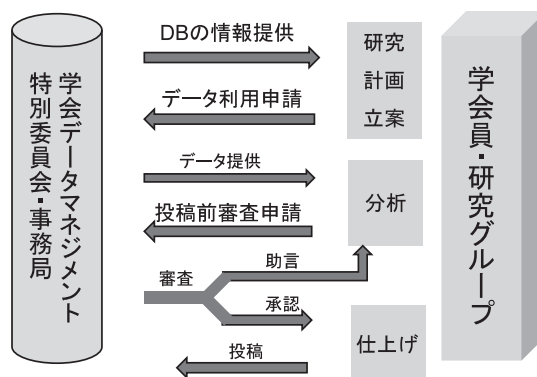


図2 データ活用のプロセス

くりである。

リハ患者データベースの登録データを活用した研究論文として、本学会誌への記念すべき第1号となるのが今号掲載の曾川論文²⁾である。学会が、データ・マネジメント・システムを開発し運用するねらいの1つは、リハ医学医療のエビデンスづくりのための研究基盤整備である¹⁾。今後もデータベースが多くの人に利用され、そこから多くの研究成果が発表されることが期待されている。そこで、小論では、リハ患者データベースの大規模データを活用する研究のプロセス（データ利用申請から分析、発表に至る流れや手続き）を紹介した後、多施設による大規模データの強み（可能性）と限界を踏まえた分析上の注意などについて述べる。

リハ患者データベース登録データを用いた研究プロセスの概要

登録データを用いた研究プロセスを、図2に示す。データベースに関する情報提供（収集）から研究計画の立案，データ利用申請，データ提供，分析，発表に向けた投稿前審査申請，審査によって（必要な場合には修正後の再審査を経て）承認を受けた後に，学会での演題や論文投稿などの形による発表に至る。

ここでは「データ利用資格とデータ入手までの手続き」，「エビデンスづくりのための分析」，「発表の事前審査」に分けて概要を述べる。以下で紹介するのは，2012年3月時点のものであり，今

後の運用経験を踏まえて見直され改良・発展していくことが見込まれる。詳細や最新情報は，リハ患者データベースのウェブサイト³⁾でご確認いただきたい。

データ利用資格とデータ入手までのプロセス

データベース開発・データ収集・入力・結合などを必要とするデータベース構築には大きな労力・コストがかかっている⁴⁾。その対価を負担せず，データベース活用による恩恵だけを甘受する者を「フリーライダー（ただ乗りする者）」などと呼ぶ。一種の公共財であるリハ患者データベースを構築・運用し，発展させるためには，労力・コストを負担しているデータ提供者が納得する「正当性」⁵⁾や「フリーライダーを許さないルール」が必要である。そのため，データ利用するには，以下のようなルールとプロセスがある。

1. データ利用資格

データを利用できるのは，原則としてデータを提供した病院の学会員のみである。ただし，学会データベースの原型となったデータバンク（厚生労働科学研究費補助金の研究班で開発）⁶⁾の入力用データベースから入力されたデータについては，同研究班における合意のもと，データ提供病院の職員（理学療法士，作業療法士，言語聴覚士など）でも活用・発表できる。また2009～2011年度は，厚生労働省の事業費や厚生労働科学研究

費補助金を受けた調査研究事業に分析・報告の面で参加協力して下さった方にも例外的にデータ提供をした。これらを除き、データ提供に伴う労力・コストを負担した会員のみがデータを利用する資格がある。

2. データに関する情報収集

リハ患者データベースに関わる情報は、ウェブサイト³⁾で情報提供している(学会のホームページの「会員のページ」→「データマネジメント事業」からリンクが貼られている)。そこには、データベース入力用ソフト(2012年3月現在、ver.2.1)から、入力基準や入力マニュアルなどデータ提供に必要な情報、小論で解説しているデータ利用プロセスにおいて必要な情報・申請書類、データベースに登録されている項目リストや寄せられた質問とそれらへの回答も掲載されている。データベース入力用ソフトは誰でもダウンロードできるので、必須(入力)項目以外に(できるだけ入力していただく)オプション項目として、どのような変数があるのか、各変数がどのような入力形式になっているのかを確認できる。今までの分析結果を載せた報告書も見ることができる。(厚生労働科学研究費補助金研究班によるデータバンクを用いた発表については同研究班のウェブサイト⁷⁾参照)

「関係者専用ページ」には、パスワードがないと入れない。パスワードは、データ提供などに協力する会員にのみ通知される。そこでは、コードブックが提供されており、それによって、前年度までに登録されたデータの各変数(必須項目)の入力件数や度数分布、配付されるデータ上の変数名やラベルとデータベースの回答選択肢などとの対応などが分かる。

3. データ利用申請と入手方法

以前から本学会には「学会が行った調査データの二次利用に関する規則」があったので、それに準拠したガイドライン⁸⁾を作成した。データ利用にはこれに沿った手続きを必要とする。データ利用の目的は、本学会の委員会活動や学術目的の利

用に限られている。データを利用した研究発表を希望する者は、様式が定められたデータ利用申請書、誓約書をダウンロードし、必要事項を記入して提出する。研究計画には、分析テーマ、背景、研究目的(仮説)、対象と方法、想定される分析結果、研究の意義などを記載する。リハ患者データベース登録データの二次利用については、データマネジメント特別委員会で審査し、分析テーマが重なる申請者がいた場合などの調整をしている。データ利用期限は、データ提供日から1年間である。データは「関係者専用ページ」からダウンロードする。このファイルにもパスワードがかかっており、利用を認められた会員にのみパスワードが通知される。

なお二次利用ガイドラインには、分析終了後に利用報告書ならびにデータをもとにした論文の報告、データ消去の義務も記されている。

エビデンスづくりのための分析

大規模データを使った研究の場合、利用できるデータの中から適切な対象群を選択することが重要である。

1. 利用できるデータ

ダウンロードできるデータは、SPSS形式とエクセル形式の2種類で、必須項目のみを含んでいる。よく使われる変数において入力ミスと思われる異常値(例えば、年齢が175歳など)については、入力した病院に照会をしたり、削除したりしたデータクリーニング後のものである。しかし、項目数は多く、登録患者数も増えているため、大規模データベースを使ったことがある者なら誰でも経験しているように、まだ疑問の余地があるデータを含んでいるであろう。分析中に異常値に気づいたら、その都度事務局に報告していただきたい。事務局から、当該病院の入力責任者に照会し修正する。

2. 分析に用いるデータの絞り込み

分析に用いるのは、研究目的に相応しい病棟・

患者グループのみである。必要に応じて分析対象とするデータを選択してから分析する。絞り込みによく使われる変数には、表1のようなものがある。

このような対象者の絞り込みをする理由は、参加施設（データ提供病院）間で、対象患者像や提供されているリハ医療に大きな違いがあるからである。そのため、単純に平均を求めると、それは典型例でなく、どこにも実在しない患者像を描く危険すらある。また、例えば歩行予後の予測をするときに、発症前から歩行不能であった患者を対象とするのは相応しくないと考えられる。

公開されているデータの二次分析では、完全な追試（結果の再現）ができるように、方法を記述することが求められる。何年に公表されたどのバージョンのデータを用いて、どの変数・条件で分析対象を絞り込んだのか、分析に用いたデータを選択プロセスを明示する。こうする理由は、同じデータベースから抽出したデータでも、対象選択の方法によって、対象患者数も得られる結果も変わってくるからである。「恣意的に患者を選択している」という誤解を避けるためにも、どのよ

表1 分析対象者の絞り込みによく使われる変数例

・直接（急性期）入院か、転入院か
・病棟種別（一般病棟、亜急性期病院、回復期リハ病棟）
・発症前の自立度やリハ開始時の重症度
・発症後入院病日（例えば、7日以内かそれ以降か）
・在院期間（30日以内、30～90日、90～180日、180日以上）など

うに選択をし、欠損値をどう処理したのかなども、図3に示すようなフローチャート図を用いて、そのプロセスを追試できるように記載する。分析上の注意については後述する。

発表の事前審査

2011年度まで、分析・学会発表を希望する会員を公募し、その演題抄録は登録前に複数の委員で査読した。それらを元にした論文についても、論文が採択されることを支援する目的で、投稿前にデータマネジメント特別委員で最低限の助言をするプロセスを導入した。これは投稿後の正式な査読とは異なり、個別の点についてのコメントまではしていないが、採択されるために最低限クリ

分析対象データ選択プロセス

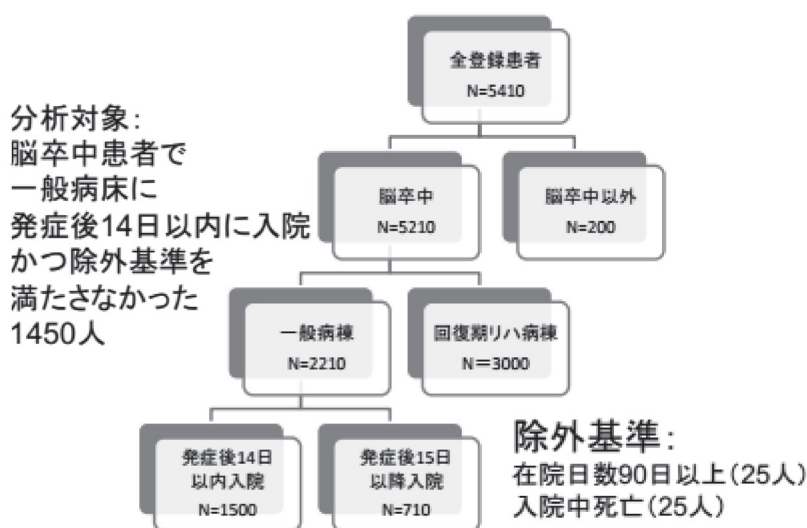


図3 分析対象データ選択プロセスのフローチャートの例

アしておくべき点（表2）がないか、ある場合にはどのような点かを、学会誌などの査読経験のある複数の委員がコメントした結果を添付して返却している。要修正と指摘された場合には、必要な再分析や論文の推敲をして論文の質を高めていただいている。

論文投稿に関わって、この間に確認されてきた点には、以下のようなものがある。①投稿雑誌は本学会誌に限定しない。②一人の研究者が作成する論文数には制限を設けず、それぞれの論文について個別に審査する。③共著者については、データマネジメント特別委員やデータ提供病院であっても、医学雑誌編集者国際委員会の統一要求項目⁹⁾における Authorship の適格性の基準に従い、1) 構想/デザイン、データ取得、データの解析/解釈などに実質的貢献をしている、2) 論文の起草、もしくは重要な知的内容に関わる批判的校閲に関わった、3) 掲載予定稿の最終承認を行っている、の3点を満たさない場合には共著者にならないで、謝辞に記すこととしている。④同じデータベースを用いた研究でも、仮説を設定しないで観察割合や関連要因を記述するようなものであれば論文種別は「調査」となるが、検証仮説を設定して新しい知見を引き出しているような論文であれば「原著」になる。⑤本学会のリハ患者データベースを用いた研究であることがわかるように方法などに記述すること、⑥分析内容・結論は学会の見解ではなく発表者の見解であることを謝辞などに明記すること、などである。

分析上の注意：強みと限界

本学会のリハ患者データベースには、多施設共同運用型の大規模データベースであることによる多くの強み（可能性）がある^{1,10)}。一方で、どのようなデータにも限界はある。それらを理解した分析と報告が求められる。ここでは分析・報告（記述）にあたって最低限注意すべき点について述べる。学会誌49巻2号掲載の「データベースマネジメント」^{1,11-13)}なども参照されたい。

1. リハ患者データベースの強み

多施設共同運用型で経年データを蓄積した大規模データには多くの強みがある。多施設のプールデータであることから、プログラムの異なる施設を対照群とした対照比較研究や施設間比較¹⁴⁾、一施設の枠を超えたより普遍的で外的妥当性¹⁵⁾の高い推論が可能となる。また経年的なデータ蓄積によって、診療報酬制度の改定前後などでの比較¹⁶⁾なども可能となる。さらに大規模データであれば、いろいろな疾患・障害像間の比較、比較的少ない疾患や障害像・合併症²⁾などについての検討、層別化や多変量解析を用いて交絡要因を調整したうえでの仮説検証、予後予測で言えば例えば半分のデータで予測式を作成し残りの半分のデータによる再現性・妥当性の検証などもできる。これらの強みを発揮した研究が期待される。

2. リハ患者データベースの限界：分析・記述上の注意

一方、登録されているデータの質に病院間の差があること¹²⁾をはじめ多くの限界も持っている。それを踏まえた分析や記述が必要であるが、いままでに審査した演題抄録や投稿原稿を見ると、考慮が不十分なものがあつた。分析・記述する上で注意していただきたい主な点を表2に示した。

1) 分析の目的にふさわしいデータを選択しているか

例えば脳卒中と大腿骨頸部骨折、一般病棟と回復期病棟とでは患者像が大きく違う。それらを一緒にしたまま平均などを求めるよりも、それらを分けて同類の患者に絞り込んで分析あるいは比較する方が望ましいことが多い。

表2 分析・記述上の留意事項

- | |
|------------------------------|
| 1. 分析の目的にふさわしいデータを選択しているか |
| 2. 分析対象データの選択プロセスを明示しているか |
| 3. 断定口調は避ける |
| 4. 関連する変数間の「逆の因果」を考慮しているか |
| 5. 「第三の因子」の影響を考慮しているか |
| 6. 「影響」「効果」という言葉を安易に使用していないか |

2) 分析対象データの選択プロセスを明示しているか

分析対象が異なれば、結果は大きく異なってくる。例えば在院日数が1～3日の者や180日を超えるような者は、どの病棟にもいるだろう。しかし、そのような患者の多くは、急な転院であったり、特別な合併症や経過があったりなど例外的な患者で、そのデータは「外れ値」を示すことがある。そのような「外れ値」を含めたまま分析すると異常な結果となる。「外れ値」を除外するためには適切な絞り込み基準が必要となる。同じバージョンのデータを使っている、例えば入院期間の基準を1～90日、あるいは7～360日にしたのでは、平均値1つとっても異なってくる。したがって、追試（再現）可能なように、分析対象の絞り込みプロセスを図3のようなフローチャートなども用いて明示する必要がある。

3) 断定口調は避ける

学会が集めているデータではあるが、日本のリハ医療の代表サンプルではない。むしろデータ提供の労力・コストを厭わなかった熱心な会員・病院の患者データだけであり、多くの病院では通年でなく8～9月退院患者など一部の患者だけが登録されている。それらによる選択バイアス（偏り）^{15,17)}は避けられず、得られた知見を安易に一般化・普遍化はできない。そのことを踏まえた考察や「今回用いた登録データに基づけば」などの限定した記述が必要である。

4) 関連する変数間の「逆の因果」を考慮しているか

「AがBをもたらす(A→B)」という因果関係があるという仮説を持って分析し、AとBの二変数間の関連を検証できたとしても、それだけでは因果関係を証明したことにはならない。関連する変数間にはほとんどの場合B→Aという「逆の因果関係」(内生性)も反映しているからである。「逆の因果」が強いために、仮説が反証されてしまう極端なケースもある。例えば、医療ソーシャルワーカーの関与(A)によって在院日数(B)が短縮することを示そうと分析してみると、期待に反して医療ソーシャルワーカーが関与した患者

の在院日数は関与しなかった患者よりも長い。これは医療ソーシャルワーカーの関与が原因でもたらされたA→Bの結果ではなく、もともと自宅退院の条件が乏しく在院日数(B)が長くなるような患者が選択されて、医療ソーシャルワーカー(A)に依頼が出されているB→Aという「逆の因果」の影響である。このような仮説と逆の結果が得られた場合には、「逆の因果」に気づかれやすいが、仮説通りの場合にも「関連性があるから」といって因果関係があるとはかぎらない¹⁷⁾ので、慎重に分析・考察する必要がある。

5) 「第三の因子」の影響を考慮しているか

分析においては、「原因」と想定している「第一の因子」と「結果」として想定している「第二の因子」とは別の「第三の因子」(交絡因子)^{15,17)}の作用を考慮する必要もある。

例えば、早期リハ(第一の因子)と退院時日常生活動作(ADL)(第二の因子)の関連を見るつもりで、発症後入院病日が3日以内の「早期リハ群」と、3日以降に始めた「非早期リハ群」を比較しても、早期リハ群に一般病棟の患者、非早期リハ群に回復期リハ病棟の患者が多い場合、あるいは一般病棟に限っても、早期リハ群にはリハが充実している病院、非早期リハ群にはリハ体制が薄い病院の患者が多い場合などに、それらを考慮しないまま分析すると誤った判断をすることになる。早期リハ(第一の因子)だけでなく、一般病棟と回復期リハ病棟の違い、あるいはリハ体制の充実度という「第三の因子」も退院時ADL(第二の因子)に影響するからである。また「早期に開始したからでなく、発症後早期の方が、慢性期よりも回復速度が速い」という時期による回復速度の違いという「第三の因子」の影響も考えなければならない。

データベースを用いて「よくデザインされた比較研究」をするためには、既知の「第三の因子」による層別化や多変量解析などを用いた分析が必要である。その場合にも、未知の「第三の因子」の影響までは考慮できていないことを自覚しなければならない。

6) 「影響」「効果」という言葉を安易に使用していないか

「逆の因果」(内生性)や「第三の因子」(交絡因子)の存在を考慮すると、無作為化臨床試験(Randomized controlled trial: RCT)や実験データでない臨床観察データの場合、いくら大規模データを使った分析でも、「因果」や「影響」を立証することは簡単ではない¹⁷⁾。ほとんどの場合、「逆の因果」や「第三の因子」の影響も含む「関連」の確認にとどまっている。

おわりに

以上、リハ患者データベースの大規模データを活用する研究のプロセスとして、データ利用申請から分析、発表に至る流れや手続きを紹介し、多施設による大規模データの強みと限界を踏まえた分析・記述上の注意などについて述べた。現在リハ医学会単独でなく、日本理学療法士協会、日本作業療法士協会、日本言語聴覚士協会との共同運用に向けて準備が進められている。その過程で、ここで紹介した内容は一部変更される可能性がある。リハ患者データベースのウェブサイト³⁾で最新情報をご確認いただきたい。

多くの会員の協力によって構築されてきたリハ患者データベースがリハ医学医療の研究基盤として活用され、今後多くの研究成果が発表されることが期待されている。

文 献

- 1) 近藤克則: データ・マネジメント・システムの概要と課題. Jpn J Rehabil Med 2012; 49: 73-78
- 2) 曾川裕一郎: 脳卒中患者における深部静脈血栓の実態調査—日本リハビリテーション医学会リハビリテーション患者データベースの分析—. Jpn J Rehabil Med 2012; 49: 137-141
- 3) 日本リハビリテーション医学会: リハビリテーション

患者データベース. Available from URL: <http://square.umin.ac.jp/jarm-db/index.html>

- 4) 近藤克則, 山口 明: エビデンスづくりに向けた大規模データベースの可能性と課題. 総合リハビリテーション 2005; 33: 1119-1124
- 5) 大久保豪, 宮田裕章, 友滝 愛, 他: 医療水準評価を目的とした大規模臨床データベースの正当性に関する研究. 医療と社会 2012; 21: 435-450
- 6) 厚生労働科学研究費補助金長寿科学総合研究事業(主任研究者 近藤克則): リハビリテーション患者データベース(DB)の開発に関する研究. 2009
- 7) リハビリテーション患者データベース(DB): Available from URL: <http://rehabd.umin.jp/>
- 8) データマネジメント特別委員会: 学会が行った調査データの二次利用ガイドライン. 2010. Available from URL: http://square.umin.ac.jp/jarm-db/download/rehabd_2jiguideline.pdf
- 9) International committee of medical journal editors: Uniform requirements for manuscripts submitted to biomedical journals. 2010. Available from URL: http://www.icmje.org/urm_main.html
- 10) 近藤克則, 山口 明, 伊勢真樹, 他: リハビリテーションにおける帰結研究—脳卒中を中心に. 大規模データベースとデータベース. 総合リハビリテーション 2008; 36: 23-27
- 11) 佐伯 寛: データベースを利用した研究の可能性と課題1—脳卒中入院患者における日常生活自立度評価の妥当性—. Jpn J Rehabil Med 2012; 49: 79-81
- 12) 宮越浩一: データベースを活用した研究の可能性と課題2—脳卒中急性期症例における合併症調査—. Jpn J Rehabil Med 2012; 49: 82-85
- 13) 田中宏太佳, 井上虎吉, 加藤文彦, 他: データベースを活用した研究の可能性と課題3—脊髄損傷データベースおよびデータベースを使用した分析の紹介: 外傷性脊髄損傷患者の職業復帰に関する因子の検討—. Jpn J Rehabil Med 2012; 49: 86-92
- 14) Jeong S, Kondo K, Shiraiishi N, et al: An evaluation of the quality of post-stroke rehabilitation in Japan. Clin Audit 2010; 20: 59-66
- 15) 名郷直樹: EBMキーワード. 中山書店, 東京, 2005
- 16) Inoue Y, Jeong S, Kondo K: Changes in hospital performance after implementation of pay-for-performance in Japan. Clin Audit 2011; 3: 17-23
- 17) Fletcher RH, Fletcher SW, Wagner EH: Clinical Epidemiology: The Essentials. 3rd Ed, Williams & Wilkins, Baltimore, 1996 (福井次夫 監訳: 臨床疫学—EBM実践のための必須知識. メディカル・サイエンス・インターナショナル, 1999)