

登録データベースシステムの設計開発及びデータ精度向上に関する検討

研究分担者 盛一 享徳 (国立成育医療研究センター小児慢性特定疾病情報室 室長)

研究要旨

【目的】令和2年度から運用開始された小児慢性特定疾病児童等登録データの二次利用申請に対し、データ抽出過程で発生した課題について検討した。

【背景】登録されているデータは、医療意見書に記載されている通りに電子化を行っているが、元となる医療意見書は、実施主体から紙媒体の写しとして登録センターに送られて来ている。登録データの二次利用申請に対応するデータ抽出過程において、医療意見書の記載の仕方に関する課題、実施主体からの紙媒体発送に関連する課題等が発生する事が予想された。

【結果】小児慢性特定疾病児童等データベースの二次利用データは、同一人物と判断された場合には、同一の共通の研究用IDを付与し、経時的にデータが追える形で提供を行っている。現在受給者番号、実名の他に5つのキー項目により、確率論的リンケージにて名寄せを行い、最終的な判断は管理者が行っている。データ抽出に際しては、データ入力ミスに依らない次の様な課題が生じた。新規申請が同一年、同一実施主体で2件あり。記載年月日は同一だが、内容が一部異なっていた。実施主体から差替前の意見書が送られてきていた。新規申請が同一年、同一実施主体で2件あり。記載年月日が離れている。内容が一部異なっている。一つは継続申請の誤りの可能性があった。全く同じ申請が2枚ある。実施主体が異なる時期に同じ意見書を送付した可能性があった。

【結論】以上のように、データ入力の誤りではなく、元となる医療意見書自体の誤りにより、複数レコードが生じるケースがあった。客観的な判断で修正が出来ない場合もあり、その際は複数レコードのままでのデータ提供とし、研究者が取捨選択できることとした。

A. 研究目的

2015(平成27)年に改正法が施行された児童福祉法(昭和22年法律第164号)に基づく小児慢性特定疾病その他の疾病にかかっていることにより長期にわたり療養を必要とする児童等の健全な育成に係る施策の推進を図るための基本的な方針(平成27年厚生労働省告示第431号)において国は、

・ 小児慢性特定疾病児童等についての臨床

データ(以下「小児慢性特定疾病児童等データ」)を収集し、小児慢性特定疾病児童等に係る医学的データベースを構築すること

- ・ 小児慢性特定疾病児童等データベースの構築及び運用に当たっては、個人情報の保護等に万全を期すこと
- ・ 小児慢性特定疾病に関する調査及び研究に有効活用できる体制に整備するとともに、指定難病患者データベースと連携することとされた。これらに基づき、小児慢性特定疾病

児童等データについては、平成 27 年の改正児童福祉法の施行以降、新しく構築されたシステムの下に、医療意見書によりデータを収集している。2020（令和 2）年より研究利用に限り、小児慢性特定疾病児童等データの提供が開始された。

登録されているデータは、紙媒体の医療意見書に記載されている通りに電子化が行われているが、登録データの二次利用申請に対応するデータ抽出過程において、医療意見書の記載の仕方や実施主体からの紙媒体発送に関連する課題等が発生する事が予想された。

本研究は、データ抽出過程で発生した課題について整理・検討することを目的とした。

B. 研究方法

令和 2 年から開始された小児慢性特定疾病児童等データの二次利用申請に係るデータ抽出作業において判明した課題を抽出した。

なお、2015（平成 27）年以降の小児慢性特定疾病児童等データの二次利用の手続きは、厚生労働省の申請窓口へ所定の様式を用いて申請を行い審査・承認された申請について、データ抽出依頼がなされる。抽出されたデータは、記録媒体に記録され、セキュリティ便にて申請窓口を經由して、申請者へ送付される。

登録レコードの名寄せ処理

小児慢性特定疾病児童等データの提供に際しては、経時的なデータ分析が可能となるよう、同一人物と判断されたレコードについては、同一の研究用 ID を付与して提供を行っている。医療意見書の記載内容の電子化を行っている登録センター内において、医療意見書の必須記載項目となっている、生年月日、性別、出生週数、出生体重、出生地の 5 つのキー項目をもとに確率論的レコードリンケージを実施しており、さらに受給者番号、氏名、登録疾病名等の情報を補完的に利用し、個体識別を行っている。

（倫理面の配慮）

本研究は、公開されている情報を元に検討を行っており、特別な倫理的配慮は必要ないものと判断した。

C. 研究結果

データ抽出の過程で、以下のような課題があることが判明した。

登録センターにおけるデータ入力に係る課題

データ抽出の過程において、以下のような課題が抽出された。

1) 医療意見書の識別番号管理

登録センターでは、全国の実施主体から送られてきた紙媒体医療意見書の写しに対し、識別番号を付与している。実施主体が誤って同一の医療意見書の写しを複数回送付するなど、何らかの理由で識別番号付与後に、登録対象から削除される場合があり、その場合は付与された識別番号の利用は停止され欠番扱いとしている。識別番号の重複チェックは登録システムにおいても実装されているが、欠番を認識する機能が無く、現在は別管理となっていることから、システム実装の不完全性が今後課題となると考えられた。

2) 必須登録項目の欠損レコードの存在

登録センターでは、データ入力のあとにダブルチェックを行い、データ登録精度を担保している。しかし登録初期において、システム実装が間に合っていなかったことを原因とした、データ入力漏れが一部存在していたことが判明した。これらは原因および原本に戻っての再入力を実施され、レコードの修正が行われた。

3) 入力済みデータの消失

データ抽出されたレコードの一部で、入力されていたはずのデータが欠損していることが判明した。原因はシステムのデータベースが Windows で実装されているサロゲートペア unicode 文字列に対応しておらず、サロゲートペアの文字が入力された段階で処理エラーとなり、入力データがクリアされて登録される、ということが判明した。

入力済みの全レコードを調査し、データ消失が起こっていたレコードを原本に戻り修正することでレコードの修復が行われた。データベース側がサロゲートペア文字列に対応できないことから、サロゲートペアの文字を入力できないよう入力端末を制限することで緊急避難的な対応となった。

4) イレギュラーな受給者番号への対応

厚生労働省より受給者番号の発番ルールが示されているが、一部の実施主体は独自のルールで受給者番号を発番しており、また同一患者であっても毎年番号を再付番される場合もある。受給者番号は二次利用提供データには含まれないが、名寄せ処理の課程ではりようされることから、これらへの対応を行った。

以上のほかに、ダブルチェックにおけるチェック漏れ、登録作業の運用を正しく反映していないシステム実装など登録作業における課題が見つかり、対応を行った。

さらに、入力作業に寄らない課題としては、以下のような点が明らかとなった。

5) 整理票の仕様における問題

登録システムの項目の設計図となっている疾患ごとのデータ整理票と呼ばれるデータ仕様書がある。データ項目には id 名が定義されているが、疾患によっては同一 id であってもデータの内容に差異がある場合がある。医療意見書開発時には、可能な限り同一項目は同一の内容に調整していたが、完全ではないことから、異なる疾患群の場合には、同一 id であっても、登録されているデータ内容が同じにならない場合がある。このため、データ提供においては、同一症例であっても疾患群ごとに分けて提供にすることとした。

6) 実施主体の医療意見書送付に係る問題

同一の人物と思われる症例において、「新規申請」が同一年、同一実施主体で複数件存在する場合があった。記載内容の一部が異なっている場合には、医療意見書の差し替えと判断し、記載年月日が新しい方を正として処理した。一方、記載内容の一部が異なっているが、記載年

月日が同じ場合は、どちらが正かの判定ができないことから、レコード重複の状態データ提供を行うこととした。また記載年月日が離れている場合には、「新規」申請ではなく「継続」申請である可能性があった。これらは実施主体から誤って複数の医療意見書が送付されてきたことに起因していた。

D. 考察

医療意見書の電子化については、単純に記載通りの内容を登録すればよいわけではなく、事前に様々な確認を行い、必要に応じて実施主体へ疑義照会を繰り返した後に実施されている。しかしながら、実施主体からの医療意見書の送付は、任意の時期に別々に送られてくることもあり、データ登録前に医療意見書の重複等を完全に排除することは難しい。また現行の登録システムは、途中で国の運用方針が変更されたこともあり、実際の運用に見合った実装が完全になされているわけではない。このため運用でカバーしている部分も多く、今後も課題抽出を繰り返しつつ対処していくこととなるだろう。

今回の検討では、データ入力の誤りではなく、元となる医療意見書自体の誤り等により、複数レコードが生じるケースがあった。客観的に判断が難しい場合もあり、その際は複数レコードのままでのデータ提供とし、研究者が取捨選択できることとした。今後もデータ精度の向上に努めつつ、研究利用をしやすいデータ提供を行いたい。

E. 研究発表

なし。

F. 知的財産権の出願・登録状況(予定を含む。)

特許取得/実用新案登録/その他
なし/なし/なし