

## 臨床効果データベースの効率的構築および運用に関する研究

研究代表者 大江和彦 東京大学医学部附属病院企画情報運営部 教授

### 研究要旨

臨床系学会等が新たに臨床症例登録 DB を構築する際には、目的、DB 利用・評価方法、IT 関連の設計・実装、登録データ項目の設定、個人情報保護や倫理指針に関する対応方法など、多様な事項を検討し関係者間で合意形成する必要がある。しかし、これらを当該臨床領域の専門家だけで事業毎に進めるには負担が大きく、短期間に効率よく DB 構築し、運用にまで持ち込むのは困難で障害となっている。そのため国内で新たに DB 事業を開始・運用するために指針となりうるガイドライン等が望まれている。本研究では臨床系学会等が新たな臨床症例登録 DB を効率的・効果的に立ち上げ運用するために必要な事項を体系的に整理し、(1) 既存 DB 事業の調査と共通課題の抽出、(2) 国外でのユーザガイドを参考にした指針作成、(3) 相互運用性と効率化の観点からの情報技術利用指針の取りまとめ、の 3 つについてこれまでの知見と利用可能な手法等に関する事例提示も含めてまとめた。

多様な症例データベースが学会や研究会主導で構築され始めている現状が把握できたが、学会や研究会などの非営利組織がデータベースを構築し運用を継続していくための人材資源と財源が不足しているとの課題が浮かび上がった。また対象症例データの悉皆性確保やデータの質確保、関連領域の横断的な連携、倫理審査とりわけ登録施設ごとの倫理審査の簡略化を求める意見が複数あった。海外のガイドラインも参考にして、患者レジストリーの構築と運営に関する手引き（総論）を提示するとともに、技術利用指針として代表的な 2 つのソフトウェアの特徴を示し、また匿名化方法の提案を行った。

### 研究分担者

植木浩二郎	国立研究開発法人 国立国際医療研究センター 糖尿病研究センター長
岡田美保子	公益財団法人 先端医療振興財団 客員上席研究員
澤 智博	帝京大学 教授

### A. 研究目的

これまで厚労省臨床効果 DB 事業により患者レジストリ DB が構築され、その他にも糖尿病症例 DB (J-DREAMS) など臨床症例 DB の構築と運用が実施されている。今後、学会等が新たに臨床症例登録 DB を構築する際には、DB 構築目的、DB 利用方針、DB 評価方法、IT 関連の設計・実装、登録デ

ータ項目の設定、個人情報保護や倫理指針に関する対応方法、症例登録基準の決定、協力医療機関での体制整備、データ入力の効率化方法、DB システムの運用管理方法、患者の追跡方法など、多様な事項を検討し関係者間で合意形成する必要がある。しかし、これらを当該臨床領域の専門家だけで事業毎に進めるには負担が大きく、短期間に効率よく DB 構築し、運用にまで持ち込むのは困難で障害となっている。

平成 27 年度には大江、植木らが「臨床効果データベースの連携及び効率的運用のための研究」において各 DB 担当者とともに各 DB の仕様の調査、各事業の課題のヒアリングを通じた課題の検討を行った。その構築基準、登録項目、運用方法、利活用と結果公表方針には個々に合目的で合理的な多様性があることが分かったが、一方で単一の DB ではなく事業間でのデータ共有により補完的なデータ活用による効率的なデータ収集や、単一 DB だけでは不可能な解析、重複登録を減らした効率的な DB 構築の実現可能性、など検討すべき多くの課題が内在していることが判明した。

平成 28 年度には、岡田、澤らが臨床効果 DB 国内外動向調査研究を実施し、海外での同様の DB 構築に際する課題、事例や体系的指針の例を調査した。本研究ではこうした事例や体系的指針を参考にして、国内で新たに DB 事業を開始・運用するために指針となりうるガイドライン等を作成する。

## B. 研究方法

### (1) 既存 DB 事業の調査と共通課題の抽出

Web アンケート調査により、これまでの臨床効果データベース事業以外の学会や研究

会が作成している患者レジストリの状況について調査した。

一方、同時期に AMED クリニカル・イノベーション・ネットワーク推進支援事業により「国内のレジストリ情報収集、利用目的毎の整理、それらの一元的把握・検索を可能にするシステム構築および公開」を目的として「CIN 構想の加速・推進を目指した レジストリ情報統合拠点の構築」（研究代表者：国立国際医療研究センター理事長 國土典宏氏）により実施された。そのため同様の調査を同じ対象者に同時期に実施することを避けるため、國土班の調査対象となりうる日本医学会分科会の学会には当研究班からの調査は行わず、國土班から調査結果の概要の提供を受けることとして連携することとした。

そこで当研究班では次のような手順で調査を行った。

#### 調査対象団体の選択：

まず、学会名鑑（日本学術会議の活動に協力する「協力学術研究団体」リスト：

<https://gakkai.jst.go.jp/gakkai/>）から臨床医学系区分の学会リストを作成（A）。

次に、UMIN の学会情報

（<http://www.umin.ac.jp/ac/>）から臨床医学系学会のリストを作成（B）。

(A)(B)を併合したリストを作成し、日本医学会加盟分科会(102 団体)を除外したリストを作成（832 団体）。

このリストから目視で臨床症例を扱う団体を選択した。（282 団体）

#### Web アンケート調査の作成：

Web 調査項目として以下の 18 問を設定した

。Web フォームの作成には Google フォームを用い、参考資料のような調査ページを作成した。

1：貴学会・団体の事業、またはそのサブグループ（研究会等）の事業として症例データベースを構築していますか？

2：登録症例の対象疾患または対象疾患領域や対象治療内容（例：胃がんの内視鏡的胃切除症例）

3：登録事業を開始した、または開始予定の時期

4：登録事業に参加している、または参加想定施設の概数

5：これまでの登録症例数の概数

6：今後の目標または想定される登録症例数の概数

7：登録症例の登録基準について登録対象期間（例：2014年から2016年の診療患者、期間限定なし、など）

8：登録症例の登録基準についての上記以外の概略（例：疾患の重症度、受けた治療内容、診療施設など）

9：症例データベースの作成目的

10：登録データの利用範囲

11：前項質問で2）または3）の選択肢を回答された場合に、登録データの利用にあたって利用条件があれば記載ください。

12：データベースシステムの開発・維持に関わる財源（複数選択可）を教えてください。

13：データベースシステムの開発・構築方法を教えてください（複数選択可）

14：システムに使用しているデータベースソフトが分かれば教えてください

15：事業の開始や運営のための公的な補助金制度があれば利用したいですか？（複数選択可）

16：1回だけ公的な補助金を利用できるとすれば、必要とする金額規模は最低で概算どれくらいですか？

17：毎年維持費を支援される公的な補助金を利用できるとすれば、必要とする金額規模は最低で概算どれくらいですか？

18：データベースを構築・運用するにあたっての課題・問題点があれば教えてください。

#### Web アンケート調査の実施：

調査は2018年3月13日に電子メールで調査依頼を282団体の事務局メールアドレスあてに発信し回答締め切りを4月末日とした。回答促進のため3月26日までに回答があった団体には謝金1000円をQuoカードで支払うこととした。

(2)国外でのユーザガイドを参考にした指針作成

岡田、澤らが平成28年度に実施した臨床効果DB国内外動向調査研究において収集した海外での同様のDB構築に際する課題、事例や体系的指針の例を参考にして、「患者レジストリの構築と運営に関する手引き」を作成した。

(3)相互運用性と効率化の観点からの情報技術利用指針の取りまとめ

大江らはこれまでに多目的臨床症例登録システムMCDRS(マックドクターズ)を開発し無償提供してきた。これを使用した

診療録直結型糖尿病症例データベース事業 J-DREAMS についてその概要をまとめ、その構築経験から知見を収集した。また代表的な症例レジストリシステムソフトウェアである REDCap について資料調査を行うことにより、これらを総合して技術利用指針をまとめた。

#### (倫理面への配慮)

本研究は実際の臨床データ等を扱わず、各臨床データベース事業の調査とディスカッションにより検討を進めたため、特段の倫理上の配慮は不要であった。

### C. 研究結果と考察

#### (1) 既存 DB 事業の調査と共通課題の抽出

##### 当研究班の調査結果：

58 団体から回答があり、20 件 (36.2%) は既存データベースあり、3 件は構築予定があり、それ以外は構築していない、または不明であった。

主要な質問に対する回答の概要を 資料 1 に示す。

1) 対象領域としては、脳血管疾患 4、耳鼻咽喉領域 3、口腔先天奇形 1、上部消化管 2、下部消化管 2、呼吸器 1、透析 1、胆道 1、産婦人科領域 1、心・循環器 3、先天性代謝異常 1、整形外科領域 1、その他 2 であった。このうち特定の治療を受けた症例を対象としているのは 7 件で 30% を占めていた。開始時期としては 1999 年以降が 23 件中 19 件であったが、1974、1987、1988 年といった 20 年以上前から構築されているものがあった。

2) 参加施設の概数については、5 から 600 台まで広く分布していたが、100-199 が 7 件と最も多く、100 施設以下が 11 件であった。登録済み症例数は、100 未満から 10 万以上まで分布し、1000-4999 例が 5 件と最多であった。

3) データベース作成目的では、患者数や患者分布の把握 16 件、疫学研究 15 件、診療への活用 8 件、自然歴調査 7 件などが上位を占めていた。一方、診療ガイドラインの検証や既存治療の有効性の検討を挙げたのは 3 件と少なかった。治験や介入研究へのリクルートもしくは治験対照群としての活用が合わせて 9 件あり、意外に多くを占めていた。

4) 登録データを利用できる範囲としては、自団体の事業に限定しているものが 15 件と多数で、自団体に所属するメンバー限定して利用を認めているものが 11 件であった。自団体以外の研究利用を認めているケースは 2 件だけであった。

5) 財源に関しては、学会自主財源、公的研究費が 15 件と大半を占め、民間企業等からの支援は 2 件しかなかった。

6) システム開発は、15 件で学会メンバーの自主開発運用がされており、業者に発注は 8 件、既製システム使用は 4 件であった。具体的に使われているデータベースソフトは、ファイルメーカー、Excel がそれぞれ 5 件、RedCAP が 3 件でそれに続いた。MCDRS の使用は 1 件であった。

7) 財源支援は、毎年の維持費の継続支援が 17 件 (74%) と最多で、最初の 1 回だけでも支援希望するものも 11 件 (48%) をしめた。

8) 課題として最も多かったのは、人材資源の不足、すなわち学会事務局でデータベースを運用する人的資源の不足とシステムや医療データのことからわかるSE（エンジニア）の不足を挙げたもので7件あった。またデータベース構築と安定的運用のための維持財源がないことを挙げるものは5件と多かった。

その他には、入力データの質維持やチェックの労力負担、登録施設側の医師の理解度の不足、セキュリティ維持、入力の悉皆性確保方法、登録医師の入力負担、データ利活用の技術的煩雑さ、登録施設での倫理審査の簡略化の必要性、データ利活用ルールの策定、領域横断的な連携、などが挙げられた。ソフトウェアシステムの機能の不足を挙げたものは2件であった。

## (2) 国外でのユーザガイドを参考にした指針作成

患者レジストリーの構築・運用に関わる代表的なガイドインとして、以下がある。

### (a) 米国厚生省 AHRQ ユーザガイド

「患者アウトカム評価のためのレジストリー- ユーザガイド 第3版」(Registries for Evaluating Patient Outcomes: A User's Guide, Third Edition) , 2012, AHRQ

米国厚生省 (Health and Human Services) の医療の研究と質に関するエージェンシー (Agency for Healthcare Research and Quality: AHRQ) が中心となり、レジストリーに関するユーザガイドを纏めている (初版2007年、第2版2010年、第3版2012年)。同ガイドはレジストリーの構築、運営にあたっての包括的ガイドである。本ユーザガイ

ドは第1巻と第2巻からなり、全体で6つのセクションからなる。第1巻は「レジストリーの構築」「レジストリーに関する法的・倫理的考察」「レジストリーの運営」の3つのセクションからなる。第2巻は「レジストリー・データと他データソースの結合に関わる技術的、法的、分析的観点からの考察」「患者レジストリーにおける特殊応用」「レジストリーの評価」からなる。

### (b) 「患者レジストリーのレジストリー」 (RoPR)

AHRQ は、臨床試験の登録システムである ClinicalTrials.gov を補完するものとして、レジストリーに固有のデータ要素を追加した RoPR (Registry of Patient Registries) システムを設計・開発している。

<https://patientregistry.ahrq.gov/>

RoPR は、レジストリー所有者の間での連携を推進し、重複を減らし、透明性を推進するためレジストリー情報を提供する。

### (c) 欧州 PARENT ガイドライン

「患者レジストリーの効率的かつ合理的ガバナンスのための方法論的ガイドラインと推奨」 PARENT (cross-border PATient REgistries iNiTiative) Methodological Guidelines and Recommendations for Efficient and Rational Governance of Patient Registries

欧州におけるレジストリーに関する動向として、PARENT ジョイント・アクションとよばれる取り組みがある。

当研究班では岡田が、海外の患者レジストリーに関するガイドラインを調査した結果

(平成 28 年度厚生労働科学研究「臨床効果に関するデータベースの国内外での動向に関する研究」[1]) に基づいて、相互利用あるいは統合的研究を妨げる要因や、共通の課題を踏まえて、患者レジストリーの構築と運営にあたって留意すべき点の概要をまとめた。これを資料 2「患者レジストリーの構築と運営に関する手引き - 総論」に示す。

本編は、特に AHRQ の「患者アウトカム評価のためのレジストリー - ユーザガイド 第 3 版」を基本としている。なお、患者レジストリーと概ね同義で使われる英語として、Patient Registry、Disease Registry、Clinical Registry、また registry を database で置き換えた用語も使われているが、以下では患者レジストリー (Patient Registry) という用語で代表する。

本編の目次構成は以下である。

1. 患者レジストリーとは
2. レジストリーの計画・設計
3. レジストリーの変更・移行あるいは終了
4. レジストリーのデータ要素
5. データソース
6. データの質・レジストリーの質
7. 倫理的・法的観点 - プライバシー・機密性
8. レジストリーと他のデータソース
9. レジストリー・データの分析・解釈・報告
10. 患者と医療提供者のリクルートとマネジメント
11. 製品安全性評価のためのレジストリー活用
12. 希少疾患のレジストリー

13. 公民パートナーシップ

14. レジストリーの所有とデータ利用

15. 維持・資金確保と利用

(3) 相互運用性と効率化の観点からの情報技術利用指針

### ソフトウェアシステムの要件

患者レジストリーデータベースを構築するためには、まず前項(2)にある「患者レジストリーの構築と運営に関する手引き - 総論」に書かれている事項を事前に十分に検討する必要がある。

次に、実際にデータベースを構築するには適切なソフトウェアシステムを準備し、データ登録環境を整える必要がある。

ソフトウェアシステムに必要な要素として

- 1) プロジェクトデザイン機能  
データベースの項目設計や入力画面・ページ遷移などを柔軟にデザイン可能
- 2) ウェブアプリケーション  
Web ブラウザだけでデザインや入力が可能
- 3) データ管理のセキュリティ要件  
データ入力・管理においてセキュアな認証・監査証跡機能を有し、利用者アカウント ID にパスワード、利用期間を個別に設定可能
- 4) 多施設研究支援  
多施設からのアクセスが可能で、多施設の複数利用者に個別に利用者アカウント ID を発行できる機能
- 5) 自立運用  
ユーザー管理をはじめ研究グループが自立して運用が可能
- 6) 柔軟性  
入力項目はフルカスタマイズが可能

7) 入力データ制約チェック  
入力データの制約（数値の条件、選択肢の制限、文字数制限、文字種制限など）を多様な形で設定できチェックできる

7) 監査証跡

ユーザーの使用履歴は追跡可能

8) データのエクスポート機能

9) 容易なインストール

10) 多言語対応

11) 国際標準に準拠

12) 電子カルテからのデータ取込などが挙げられる。

こうした機能のある程度有するシステムとして REDCap、MCDRS が挙げられる。

REDCap (Research Electronic Data Capture) は、臨床研究に関するデータ収集とデータベース管理を目的として開発されたウェブアプリケーションである。多様なデータの種類についてサポートしていると同時にセキュリティ面での対策がなされ、オンラインあるいはオフラインのデータキャプチャを支援するように設計されている。REDCap に関する支援団体である、REDCap コンソーシアムが構成されており、100以上の国々から数千の施設が参加している。詳細を資料3に報告する。

MCDRSは本研究代表者がこれまでの開発してきたものである。MCDRS(マックドクターズ)はWeb ベースの臨床症例データ登録システム用のソフトウェアで、研究施設や学会団体等でMCDRSをサーバコンピュータにセットアップすると、そこに所属する研究チームのDB 担当者(研究DB 管理者)は症例データベース画面をWeb 上でいくつでも作成できる。研究DB 管理者は症例登録する利用者にID とパスワードを発行することで、その利用者はWeb 上でこの登録画面にアクセスして症例登録ができるよ

うになる。作成したDB は研究DB 管理者がダウンロードしてエクセル、統計ソフトで集計、解析できます。

MCDRS の大きな特徴としてSS-MIX2標準化ストレージが設置されている医療機関からデータ登録する場合には、一定の設定をすればそこからデータ転記ができる点が挙げられる。症例データの登録は医療者にとって手間のかかる作業であり、特に電子カルテに入っているデータを見ながら手入力していくことは大変苦痛な作業で、間違いも起こりやすい作業である。もし自病院にSS-MIX2 標準ストレージと連携ゲートウェイが設置されていれば、そこから取得できる患者基本データ、検査データ、処方データなどを少ないクリックでWeb 画面上のデータ入力欄に自動転記できる仕組みがMCDRS には装備されている。この機能のおかげで、これらのデータを一瞬にして正確にWeb 画面上に転記でき、登録ボタンを押すだけで症例登録ができるようになる。ただし施設の医療情報管理者やネットワーク管理者との十分な調整と管理者による事前のシステム設定などが必要である。

MCDRSについての詳細は <http://mcdrs.jp> から情報を得ることが出来る。

MCDRSを活用した糖尿病症例登録データベースとして、診療録直結型全国糖尿病データベース事業(J-DREAMS)が挙げられる。

資料4にJ-DREAMS事業の概要をまとめた。

### 匿名化(仮名化)の方法

症例登録では登録施設での患者番号(カルテ番号)情報を直接登録せず仮名化して登録することが望まれる。一方で、仮名化後の番号と元の患者番号との関係は登録施設では対応表として管理される必要がある。

改正個人情報保護法ではデータの出し元ルールに基づき対応表を有する症例データの登録は個人情報、要配慮個人情報の提供となる。しかし、そもそも症例登録の目的や利活用が学会等が実施する学術利用であれば、個人情報保護法第4章の適用除外となるため、個別のオプトイン同意が不要であり、オプトアウトによる同意のみで運用可能である。

その場合でも仮名化は重要であり、その方法として医療機関での患者番号、性別、氏名文字列、生年月日文字列などを連結してハッシュ関数によりハッシュ値を求めてそれを仮名化IDとして利用することが、よくおこなわれるようになってきた。ハッシュ関数によりハッシュ値の算出方法として良く使用されるアルゴリズムはSHA-256という方法である。WebIT 辞典

(<http://e-words.jp/w/SHA-256.html>) から引用すると「SHA-256はハッシュ関数の計算手順(アルゴリズム)を定義しており、最長で2の64乗-1ビットまでの任意の長さのデータから256ビット(32バイト)のハッシュ値を生成する。同じ原文からは必ず同じ値が得られる一方、少しでも異なる原文からはまったく違う値が得られる。いわゆる暗号的ハッシュ関数として設計されており、ある原文を元に同じ値になる別の原文を効率よく探索することは困難なようになっている。SHA-256はNSA(米国家安全保障局)が考案し、2001年にNIST(米国標準技術局)によって連邦情報処理標準の一つ(FIPS 180-4)として標準化された「SHA-2」規格の一部として定義されている。SHA-2では他にハッシュ値の長さが224ビットのSHA-224、384ビットのSHA-384、512ビット

のSHA-512などが定義されており、SHA-256はその中でも実装のしやすさや計算速度、暗号的な安全性のバランスに優れ、最も広く普及している。」とある。

前述のMCDRSでは、設定により症例登録時に自動的にSHA-256によるハッシュ値を生成して患者IDの代わりに登録するとともに、患者IDとハッシュ値との対応表を自動的に施設内データとして保存する機能がある。

異なるデータベース事業間で同一のハッシュ値を用いることは個人情報保護法ガイドライン上は推奨されないこととされていることに留意が必要であろう。

#### D. 結論

多様な症例データベースが学会や研究会主導で構築され始めている現状が把握できたが、学会や研究会などの非営利組織がデータベースを構築し運用を継続していくための人材資源と財源が不足しているとの課題が浮かび上がった。また対象症例データの悉皆性確保やデータの質確保、関連領域の横断的な連携、倫理審査とりわけ登録施設ごとの倫理審査の簡略化を求める意見が複数あった。

本研究班では、海外のガイドラインも参考にして、患者レジストリーの構築と運営に関する手引き(総論)を提示するとともに、技術利用指針として代表的な2つのソフトウェアの特徴を示し、また匿名化方法の提案を行った。

#### E. 健康危険情報

なし



## F. 研究発表

### 1. 論文発表

1. 大江和彦,新しい医療技術 医療における ICT 活用のイノベーション,整形・災害外科 61(1), 91-95, 2017
2. Takehiro Sugiyama, Kengo Miyo, Tetsuro Tsujimoto, Ryota Kominami, Hiroshi Ohtsu, Mitsuru Ohsugi, Kayo Waki, Takashi Noguchi, Kazuhiko Ohe, Takashi Kadowaki, Masato Kasuga, Kohjiro Ueki, Hiroshi Kajio: Design of and rationale for the Japan Diabetes compREhensive database project based on an Advanced electronic Medical record System (J-DREAMS). Diabetology International. 8(4), 375-382, 2017 Jun 27.

### 2. 学会発表

1. 医療ビッグデータの活用と展望,口頭,大江和彦,医療ビッグデータの将来展望と臨床研究, 2018/1/31,東京
2. ビッグデータと AI がもたらすこれからの医療,口頭,大江和彦,エジソンの会 第 18 回会合, 2018/2/20,奈良
- 3 ビッグデータと AI で医療は変わるだろうか,口頭,大江和彦,次世代産業ナビゲーターフォーラム, 2018/3/13,東京

4. IoT・ビッグデータ・AI がもたらすこれからの医療,口頭,大江和彦,Medical

Information Conference in Kawasaki, 2018/3/14,川崎市

5. ビッグデータと AI 時代の医療,口頭,大江和彦,研修医・若手医師キャリアデザインの会～糖尿病学へのいざない～, 2018/3/27,東京

6. SS-MIX2 活用の動向,口頭,大江和彦, SS-MIX2 シンポジウム 2017, 2017/8/29,東京

7. 電子カルテ情報を活用した多施設臨床効果データベースの構築と課題,ポスター, 桑原篤憲,片岡浩己,柏原直樹,岡田美保子,J-CKD-DB 運営企画委員会,第 37 回医療情報学連合大会,2017/11/22,大阪

## G. 知的財産権の出願・登録状況

### 1. 特許取得

なし

### 2. 実用新案登録

なし

### 3. その他

なし