

データの管理

<教材提供>

AMED 支援「国際誌プロジェクト」提供

無断転載を禁じます

無断転載禁止

目次

はじめに

データ収集までの準備(臨床研究を開始する前)

事前計画

症例報告書

データ管理システム

データ収集の実際と管理(臨床研究実施中)

同意の取得について

個人情報の取り扱い

データの収集・保管方法

データマネジメント

データのコーディング

データの修正と固定

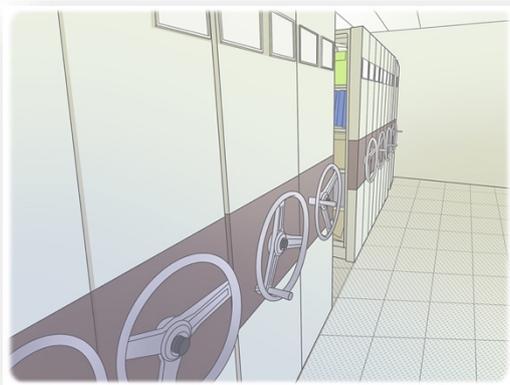
データの保存(臨床研究終了後)

まとめ

参考文献

はじめに

臨床研究から信頼性の高い成果を得るためには、その臨床研究が適切に計画・実施されていることが保証されている必要があります。臨床研究の信頼性を保証する方法の一つとして、人を対象とする医学系研究に関する倫理指針(平成29年2月28日一部改正、以下「倫理指針」)では、侵襲を伴う介入研究についてモニタリングや監査を行うことを研究責任者の責務としています。研究の進捗や組織体制についての保証だけでなく、収集されるデータに入力の誤りや入力者によるばらつきが混入されることを避け、正しいデータ収集が可能になるように努力する業務が**データマネジメント**です[1]。第三者によって調査結果を検証できるということが、データの信頼性を保証することにつながります。



学習目標

- ・全研究期間を通じたデータマネジメントとデータの管理について理解する。

データ収集までの準備(臨床研究を開始する前)

事前計画

データマネジメントの話をする前に、臨床研究における「統計解析」はどの段階に行う作業か考えてみましょう。「さあ、解析だ！」と、統計ソフトにデータを読み込んで検定を行うのは、データの収集が終わり、データ固定(データのクリーニングがすんでそれ以降はデータの修正を行わないこと)の後です。しかし、データ固定に至るよりはるかに前の研究立案時から、「統計解析」は意識しておく必要があります。統計解析の方法は、データ収集が終わった後に集まったデータを眺めながら決めるものではありません。研究計画段階で、当該研究によって明らかにしたいこと、そしてそのために検証すべき内容(主要・副次評価項目)が決まった段階で決定します。そして、その統計解析を実施する上で必要なデータを明確にして初めて、収集項



目(データ)が決定できます。解析計画は事前に明記されるべきであり[2]、研究計画書に記載することが求められています(倫理指針、臨床研究法)。もちろん、長期のコホート研究やレジストリー研究など、具体的な評価項目が明確に決定されないまま開始される研究もありますが、統計解析を意識せずにやみくもにデータを集めると、解析に必要な項目が収集できていなかったということになりかねず注意が必要です。

時々、研究計画書の収集項目の欄に「血液検査」とだけ書かれていたり、「バイタルサイン: 血圧、脈拍、など」と記載されていたりするものを見かけますが、このような曖昧な記載ではデータを正確に収集することは困難です。臨床研究は一人で何もかもやれるわけではなく、関連業務を適切に分担して遂行されますので、研究分担者や協力者が判断に迷うような記載は避けなければなりません。

症例報告書

どのような項目を収集するかが決まったら、「**症例報告書**」を作成します。症例報告書とは研究計画書で規定された臨床データを収集するための記録用紙です。CRF (Case Report Form あるいは Case Record Form) や、DCF (Data Collection Form) など、さまざまな名称で呼ばれることがあります。データを収集するためだけでなく、モニタリングを補助する役割もあります[1]。CRF は、臨床研究ではデータマネジャーが、研究者や生物統計家の意見を聞きながら作成します。例えば「血圧」のデータを収集するとしましょう。収縮期血圧なのか拡張期血圧なのか、平均血圧なのか、その計算式はどのようなものか、臥位で測定するのか坐位で測定するのか、1回測定なのか2回測定した平均値を収集するのか、左右どちらの腕で測定するのか、横断研究であれば1回の収集でいいでしょうか、前向き研究の場合は、登録時のみでいいのか、毎回の受診ごとに必要なのか...など、実際に現場でデータを収集する人にとっては考えるべきことがたくさんあります。そこで、すべてのデータについてどのように収集すべきかがわかるように、研究計画書および症例報告書を作成することが重要です。

研究計画書に規定されている統計解析に必要な情報が、症例報告書にすべて網羅されていることは言うまでもありません。同時に、無駄が少なく不必要なデータがないという点で効率的なものでなければなりません。せっかく収集するのだから、とあれもこれも、となりがちですが収集しても使わないデータは収集すべきではありません。ほかにも症例報告書に必要な要件について、表1にまとめました(参考文献 1, p71 より引用改変)。

データ管理システム

収集された症例報告書は、データセンターで一つのデータセットとして統合されます。研究者主導の臨床研究では、データセンターを特には設けずに、Excel などのスプレッドシートソフトを使って、1 行目に変数名、1 列目に被験者の ID が来るようなデータベースを作成し、直接データ集積を行っていく方法も多く用いられています。近年では、症例報告書を電子的に作成 (eCRF) し、インターネット上にデータベースを構築する Electronic Data Capture (EDC) が多く用いられるようになってきました。EDC を用いた研究では、インターネットに接続可能なコンピュータ端末さえあれば、各参加施設の診察室からデータベースにアクセスし、データを入力することができます。EDC の利点としては、入力者および閲覧者管理が可能なこと、データ参照・データ入力・データ修正のそれぞれについて監査証跡が残せること、入力規則の設定が可能であり誤入力や欠損を防げることなどが挙げられます (表 2)。

EDC の構築に必要な費用は高額というイメージを持つ研究者も多いかもしれませんが、UMIN-INDICE[4, 5]や REDCap[6, 7]などは比較的安価に利用することが可能です。詳細は他項に譲りますが、EDC は、モニタリングやデータチェックなどの効率をあげることに貢献します。つまり、データの精度をあげ、研究全体の質の向上が期待できます。

データ収集の実際と管理 (臨床研究実施中)

同意の取得について

平成 29 年に改正された個人情報保護法等 (個人情報の保護に関する法律: 個人情報法、行政機関の保有する個人情報の保護に関する法律: 行個法、独立行政法人等の保有する個人情報の保護に関する法律: 独個法) では、病歴などの情報は「要配慮個人情報」として扱われ本人の同意のない収集は認められていません。しかし、病歴情報が共有できないと臨床研究が行えません。倫理指針では、人体から取得した試料 (血液、体液、組織、細胞、排泄物及びこれらから抽出した DNA 等、人の体の一部であって研究に用いられるもの (死者に係るものを含む。))



を用いない研究であれば、インフォームド・コンセントを必ずしも要しないものの、インフォームド・コンセントを受けない場合には、原則として研究対象者等の適切な同意を受けなければならない、としています。

「適切な同意」の中には、研究対象者等からの直接の同意取得がかなわない場合のオプトアウトも含まれています。オプトアウトとは研究の目的や方法、収集項目などの詳細を公開または通知することによって、「研究対象者等が拒否できる機会を保障する」ことです。

このほか、倫理指針には収集するデータが既存のものなのかこれから収集されるのか、匿名化されているか否かなどによって、データを利用する場合の取り扱いを定めています。前向きに収集されるデータであれば、研究対象者から研究参加についての同意のみでなく、データを研究目的で使用するについての同意を得なければなりません。また、通常の診療で得られた臨床データは、当初は研究目的で収集していませんので、こうした既存データを研究に再利用したい場合も同様の手続きが必要です。特に臨床研究法が施行された今日では、これらのルールについて、知らなかったではすまされません。倫理指針や臨床研究法の説明は本項の目的外であるため割愛しますが、研究を実施する前には内容を十分に理解しておくことが重要です。

個人情報の取り扱い

臨床研究のデータを収集するにあたって最も注意すべきことは、個人情報の取り扱いです。倫理指針のガイドラインには、「個人情報」「個人情報等」「個人に関する情報」「個人識別符号」「要配慮個人情報」それぞれについての定義が記されています。これらの情報を扱う際に必要な行為として「匿名化」という方法があります。臨床研究では、個人情報や個人識別符号などを削除して、個人とかわりのない何らかの記述に置き換えることによって、研究対象者を識別できるようにします。「症例登録番号」や「症例識別番号」などと呼ばれるこの記述(数字だけのこともありますし、数字とアルファベットを組み合わせることもあります)は、その研究の中で一意の値や記号で表し、症例報告書の記載時にも使用します。症例登録番号が誰のものなのかを照合できるようにしたものを「対応表」と呼びます。対応表をデータセットとは別に保管することによって、研究データから個人情報が漏洩することを防ぐことができます。Excel ファイルにパスワードをかけるから研究対象者氏名を入力しても大丈夫だと思っている研究者もいます

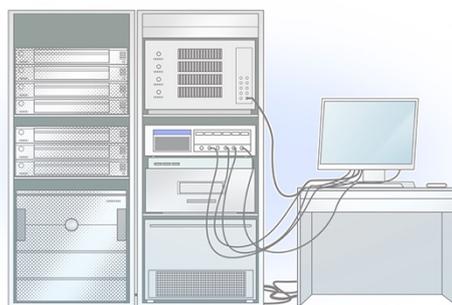


が、これらの扱いについても指針で記載されており、研究計画書にも管理方法を明記しておく必要があります。

データの収集・保管方法

データ収集の具体的な方法については、前述のとおり紙の症例報告書や Excel などのスプレッドシートを用いる方法や EDC の利用などが挙げられます。データ集積期間中に注意すべき点は、施設間におけるデータ授受の方法とデータセンター（研究代表者や事務局のこともあるでしょう）における保管の方法です。まず、EDC を用いる場合であれば、各参加施設からデータベースに直接アクセスすることが可能ですので、入力されたデータはインターネット等を通じて、サーバーに蓄積されます。では、紙の症例報告書や Excel ファイルはどのように授受することが望ましいのでしょうか。これらの中に個人情報が含まれていないことは大前提ですが、例えばデータの入力された Excel ファイルをメールで送受信することは、誤送信によるデータ流出の恐れがありますので、たとえ前述のような匿名化をしていたりパスワードをかけていたとしても避けましょう。USB メモリーや CD-R などの電子媒体へコピーし、送受信記録を残した上で、郵送による授受を推奨します。

治験や臨床研究におけるデータセンター業務を専門としている機関では、データ管理システムの利用について厳しい管理条件があります。メンテナンスやバリデーション、データのバックアップなど GCP[2]や ER/ES 指針[8]などでも細かく規定されていますが、ここでは研究者自身でデータの収集・管理を行う場合のデータの保管について、特に、USB メモリーや CD-R、または紙でデータが収集された際の保管方法について考えてみましょう。いうまでもなく個人のノートパソコン 1 台にデータを入れて持ち歩くのは推奨できません。可能であれば、鍵のかかる研究室や医局に設置された持ち出しできないコンピュータのハードディスク上で管理し、USB メモリーや CD-R は、同じく鍵のかかるキャビネットにしまっておきましょう。紙で集められたデータは、Excel などで作成したデータセットに入力した上で、同じく鍵のかかるキャビネットに保管します。モニタリングの章に詳細がありますが、紙の CRF は原資料となりますので、廃棄してはいけません。倫理指針のガイダンスでも、物理的安全管理と技術的安全管理という言葉を用いて詳細に説明されています。これらの管理方法については、研究開始前にきちんと検討し、研究計画書に記載する必要があります。



データマネジメント

臨床研究データを収集するにあたって、最終的に統計解析に用いる際に適切に取り扱うことができるよう意識して、データセットを作ることが重要です。データの収集が始まったら、データセットに対し正確性のチェックが可能になります。症例ごとに行うのか、登録時データが揃ったときにまとめて行うのか、どれくらいの頻度で行うのか、などについてはその研究の品質管理計画に従い、データマネジメント計画書あるいはモニタリング計画書で規定されます。

データのチェックを行う際には、割付結果やほかのデータとの相関などの情報を可能な限り考慮しない状態で、データの正確性について得られたデータだけを見ながらチェックしていきます。盲検下で実施された試験であれば、これを「盲検下レビュー」と呼び、割付結果が開示される前に行うチェックとして位置づけられています。また、非盲検試験や観察研究でも実施できる簡単なチェックとしてダブルエントリーという方法もあります。2回のデータ入力を行う方法であり、1回目と2回目のそれぞれの入力内容を比較することにより入力ミスを発見する方法です。1名で2回入力する場合がありますし、2名が1回ずつ入力する場合があります。その他、2名で読み上げながらのデータチェックという方法(マニュアルチェック)もありますが、実際に統計ソフトにデータを読み込んで、欠損値の有無をチェックしたり、箱ひげ図などを書くことによって、外れ値を効率的にチェックする方法もあります。

データマネジメントの担当者(データマネジャー)は、データチェックで抽出された不適切データを入力者や各施設の担当者に差し戻し、確認や修正を依頼します。データセンターから発信される問合せ作業を「クエリ発行」と呼びます。データマネジャーから直接発行される場合もありますし、モニタリング担当者を介することもあります。他の職種との連携が必要不可欠であることから、コミュニケーション能力を要しますし、時には医学的な知識が必要になることもあります。日本の医療現場においては、臨床研究におけるデータマネジャーという役割はまだ一般的でないかもしれませんが、統計解析を理解し、データの信頼性および研究全体の質を向上させるために、今後さらにその重要性が増す職種といえるでしょう。

データのコーディング

臨床研究データの中には、そのまま用いるよりも何らかの値でコード化されていたほうが集計しやすくなるデータがあります。例えば、合併症の有無はテキストデータで取るよりも、「有」が 1、「無」が 0 などというように、データセットには数値で入力されていたほうが統計解析の際に取り扱いやすいことがあります。また、合併症の疾患名も、テキストデータで入力したままの状態では集計が困難です。



データに対して特定のコードを割り当てる作業をコーディングといいます。コーディングには、データマネジャーが独自に作成するコード値定義表によるコード化（有無の有=1、無=0 など）のほか、辞書を利用したコード化（シソーラス thesaurus といいます）もあります。データマネジャーが当該臨床研究に合わせた入力ルールを策定し、それに沿って管理することで、はじめて整然とコード化されたデータとなります。多くの場合、コード値定義表で規定されたコードをもとに、データをコンピュータに入力する際にコーディングが行われますが、EDC の集計機能や辞書をうまく利用すれば、データを記入された通りにコンピュータに入力しておき、データ固定前にコードを割り当てることも可能です。辞書に関して言えば、疾患名や有害事象の症状は、MedDRA (Medical Dictionary for Regulatory Activities Terminology) という ICH で合意された辞書に基づいてコード化するのが通常の方法ですが、研究実施中に辞書がバージョンアップされた場合の取り扱いなどを最初に決めておくなどの注意が必要です。バージョンアップは半年に 1 回程度なされており、2018 年 9 月時点でバージョン 21.1 です。併用薬については、WHO-DD (WHO Drug Dictionary) や医療用医薬品名データファイル等の辞書があります。

データの修正と固定

紙の CRF においてデータの訂正を行う場合には、訂正箇所にも二重線をひいて、日付、訂正理由の記載、署名・捺印を行い、改ざんでないことが記録として残るようにします。電子データの変更については、前述のとおり EDC に修正履歴を残す機能が搭載されていれば問題ありませんが、修正履歴を残す機能が搭載されていないソフトウェア (Excel 等) でデータセットを作成している場合には、別途、データマネジメント担当者自身が、修正記録を作成しておく必要があります。修正後のファイルで完全に上書きしてしまうのではなく、修正前のファイルのコピーを別名保存するな

どしておきましょう。

それぞれの症例データについての問題点や疑問点が解決された状態を「症例固定」と呼び、データベース上でもそれ以上の変更が行われないという段階を「データ固定」と呼びます。データ固定が行われたデータが、統計解析の対象となります。データ固定前に統計解析を行ってしまうと、解析結果を見てからデータを修正した可能性を否定できなくなり、研究の信頼性を低下させることとなります。統計解析の過程の中で、データの再確認が必要になったり、追加収集が必要になったりしないよう、収集するデータの項目や頻度、収集の方法などについて計画段階で十分に検討することが重要になります。

データの保存(臨床研究終了後)

研究期間中のデータの保管については、前述のとおりですが、臨床研究が完了した後も、紙資料、電子データなどすぐに廃棄してはいけません。倫理指針では、「第 8 章 研究の信頼性確保」に「第 20 研究に係る試料及び情報等の保管」として、「可能な限り長期間保管されるように努めなければならない」としています。また、侵襲を伴う介入研究では、研究機関の長は研究終了(を報告した)日から 5 年を経過した日又は当該研究の結果の最終の公表、つまり最後の論文が公表されてから 3 年を経過した日のいずれか遅い日までの期間、適切に保管されるよう必要な監督を行うこととしています。試料や情報の授受に関する記録を残すことについては前述のとおりですが、ここで作成された記録についても、試料や情報を提供する側は提供した日から 3 年、提供を受ける場合は研究終了が報告された日から 5 年は保管すると期限を定めています。5 年経過したら削除すべき、ということではありません。今後の医学研究にとっても貴重な情報となり得るデータですので、将来、別の臨床研究のデータの一部として解析を行うということも可能です(二次利用)。いずれにしても、廃棄されるまでの期間、適切な方法で保存されていることが重要です。



まとめ

データ管理を適切に行うことは、臨床研究の質を保証する上で重要です。故意ではなかったとしても、データの修正が改ざんとみなされてしまったり、被験者の情報が漏洩してしまったりしては、研究の質を低下させるだけでなく、世界の中での日本の臨床研究の信用を失墜させることにもなりかねません。関連する法令や指針を理解し、どんな小さな研究であっても、精度の高いデータからエビデンスを発信できるよう、研究者一人一人が心がけておく必要があります。

表 1 症例報告書の作成にあたって満たすべき要件(参考文献 1、p.71 より引用改変)

-
- ・ 研究計画書に規定された項目がすべて揃っていること
 - ・ 無駄が少なく効率的であること
 - ・ 理解しやすく、間違えにくいこと
 - ・ 記入しやすい／入力しやすいこと
 - ・ 使用するコンピューターデータベースとの整合性がとれていること
 - ・ 集計・解析しやすいこと
 - ・ 保存性に問題がないこと
-

表 2 EDC を用いる利点(参考文献 3、p.284 より引用)

1. データが直接データベースに入力されるため、紙媒体で集めてから入力する場合に比べると、入力ミスを減らすことができる。
2. 入力画面に、データチェック機能をプログラムしておけば、異常な値が入力された場合にすぐ対象者にそれがフィードバックされる。
3. 入力システムに飛び先指定のロジックを組み込むことができ、たとえば、タバコの本数を聞く質問は、喫煙経験を問う質問に「あり」と答えた人しか答えられないようにすることができる。
4. 画面入力、研究スタッフだけではなく、本人が直接入力することができる。本人が直接入力する場合には、個別面接の場合に比べ、性行動や非合法ドラッグの使用といった微妙な内容についても答えやすい。

5. ノートパソコンや携帯情報端末などの携帯が可能で無線送信ができる装置を用いて、さまざまな場所での入力ができる。

本単元は、日本医療研究開発機構(AMED)が実施する研究公正高度化モデル開発支援事業(第一期)の「医系国際誌が規範とする研究の信頼性にかかる倫理教育プログラム」(略称: AMED 支援国際誌プロジェクト、信州大学・大阪市立大学)によって作成された教材です。作成および査読等に参加した専門家の方々の氏名は、[こちら](#)に掲載されています。

無断転載禁止

参考文献

- [1] 辻井敦著、大橋 靖雄監修 (2004). 臨床試験データマネジメント-データ管理の役割と重要性、医学書院、東京
- [2] ICH-E6 GCP (医薬品の臨床試験の実施基準)、<http://www.pmda.go.jp/int-activities/int-harmony/ich/0028.html> (最終閲覧日 2019 年 3 月 7 日)
- [3] 木原雅子、木原正博訳(2009). 医学的研究のデザイン第 3 版 研究の質を高める疫学的アプローチ、メディカル・サイエンス・インターナショナル、東京
- [4] UMIN インターネット医学研究データセンター、<http://indice.umin.ac.jp/> (最終閲覧日 2019 年 3 月 7 日)
- [5] 曾根美雪、荒井保明、木内貴弘、石川ひろの、青木則明、稲葉吉隆、吉岡哲也、新槇剛、小林健、松岡利幸、穴井洋、谷川昇、大須賀慶悟、竹内義人、奥坂拓志、金澤右、松井修、遠藤啓吾(2012). 医師主導の多施設共同臨床試験における UMIN インターネット症例登録センター (UMIN-INDICE) の活用 : 日本腫瘍 IVR 研究グループ (Japan Interventional Radiology in Oncology Study Group: JIVROSG) での評価. 癌と化学療法 39(4): 619-623
- [6] REDCap、<https://www.project-redcap.org/> (最終閲覧日 2019 年 3 月 7 日)
- [7] 太田恵子、山本景一、榊原恵、甲斐陽子、高橋佳苗、原口亮、桑田成規、山本晴子(2015). 研究者主導臨床研究向け無償 Electronic Data Capture システム「REDCap」を用いた Computer System Validation プロジェクトの事例報告. 薬理と治療 43(suppl-1): 5072-5084.
- [8] 厚生労働省医薬食品局長「医薬品等の承認又は許可等に係る申請等における電磁的記録及び電子署名の利用について」(薬食発第 0401022 号)
<https://www.pmda.go.jp/review-services/inspections/gcp/0007.html> (最終閲覧日 2019 年 3 月 7 日)