

公募シンポジウム3

国際的医薬品規制・臨床研究におけるリアルワールドデータ活用のいま

2020年11月19日(木) 16:20 ～ 18:00 F会場 (イベントホール・特設会場1)

[2-F-3-03] 医薬品規制領域におけるデータの標準化動向：相互運用性・実装の課題と今後

*岡田 美保子¹（1. 一般社団法人医療データ活用基盤整備機構）

*Mihoko Okada¹（1. 一般社団法人医療データ活用基盤整備機構）

キーワード：real world data, data interoperability, data quality, common data model, pharmaceutical affairs

RWD(Real World Data)については国際的に合意された、ただ1つの定義が存在するわけではないが、およそ電子カルテや診療報酬データ等、日々の医療で生成・収集されるデータが中心であり、加えてセルフモニタリングデバイス等から収集されるデータ、患者レジストリー・データ等も含まれる。ICHのGCP Renovationペーパー(Jan 2017)では、近年のリアルワールドから得られる大量データの活用の可能性が言及されるとともに、各種リアルワールドソースから得られるデータについては国際的に合意された標準がないことが大きな障害となっていると指摘されている。RWD/RWE(Real World Evidence)に関わる標準の議論は、ヘルスケアと臨床研究・規制領域にわたることが1つの特徴であると考えられる。ヘルスケア全般を対象とするHL7と、臨床試験を対象とするCDISCには元より連携があるが、近年はCDISCにおけるHL7 FHIRの取り組みがあり、また欧州EMAを中心とする医薬品関係のFHIRリソース開発などがある。また、データ分析に用いるデータベースのためのCDM(Common Data Model)も1つの特徴である。RWEに関する論文に関して、現在「データの質」の問題がクローズアップされている。データの標準とデータの質は、どう関わるのか、これにはデータの質のいかなる点を論じるかを明確にする必要がある。本報告では「データの質」として、1) 要求水準への適合性と、2) 利用目的に照らしての適合性を考え、RWD/RWEで求められるRWDの質向上と、OMOP-CDM、PhUSE、BRIDG、CDMH(Common Data Model Harmonization)等の取り組みに言及しつつデータ標準が目指す「データ interoperability」について考えてみたい。

医薬品規制におけるデータの標準化動向

- 相互運用性・実装の課題と今後 -

岡田 美保子^{*1}

^{*1} 一般社団法人医療データ活用基盤整備機構

Data Standard Trends in Clinical Studies and Pharmaceutical Affairs

- Data interoperability: issues and the way forward -

Mihoko Okada^{*1}

^{*1} Institute of Health Data Infrastructure or All

Sources for healthcare Real-World Data (RWD) are generally considered as routinely generated and collected clinical data, including Electronic Health Records (EHRs), administrative or claims data, data from self-monitoring devices, and so on, although there is no consensus on a definition of RWD. There is an increasing expectation to derive reliable Real-World Evidence (RWE) using RWD, although it is still unclear whether observational data can be used to address the same clinical questions being answered by traditional RCT. In the use of RWD for research, two major issues are data standards and data quality. It should be noted that there are no universally accepted standards currently in use for these different real-world sources, and this is considered an impediment to the use of RWD for RWE. There have been efforts to improve the current state by developing standards to overcome the challenges confronted. The discussions on data standards for RWD/RWE extend across healthcare and clinical research/regulatory domain, and the issues of data quality gain prominent attention. How the data standards impact the data quality depends on what aspects of quality we are concerned about. We will discuss on-going efforts to develop standards for making clinical data interoperable for clinical research and for regulatory use, and challenges confronted including quality assurance of data.

Keywords: real world data, data interoperability, data quality, common data model, pharmaceutical affairs

1.はじめに

近年、医薬品規制領域において、リアルワールドデータ (Real World Data: RWD)を用いたリアルワールドエビデンス (Real World Evidence: RWE)の創出が注目されている。臨床研究における RWD の適用、RWE の創出に関しては、国内外で様々な観点から取り組みがなされている。

RWD については、「データ標準」と「データ品質」の二つの課題がクローズアップされている。過去、ヘルスケア領域と臨床研究・薬事領域におけるデータの議論は、それぞれの領域で分かれていたが、RWD/RWE においては両領域にわたる取り組みが求められることが、1 つの特徴である。ここでは、臨床研究、医薬品規制領域における RWD 活用の動向を概観し、データ interoperability を中心に、本領域に求められる標準について考えてみたい。なお、以下における日本語訳は暫定的であり、異なる英単語に同じ和訳が用いられる場合や、同じ英語でもコンテキストにより和訳が異なる場合もあるため、以下では、なるべく原文の英語表記も示している。

2.医薬品規制領域における RWD の議論

ICH の GCP Renovation ペーパー¹⁾では近年のリアルワールドから得られる大量データの活用可能性が言及される一方で、各種のリアルワールドソースから得られるデータについては国際的に合意された標準がないことが大きな障害となっていることが指摘されている。リアルワールドデータ (RWD)という用語については国際的に合意された、ただ 1 つの定義が存在するわけではないが、Framework for FDA's Real-World Evidence Program²⁾ (以下 FDA RWE のフレームワーク) に述べられるように、およそ EHR (Electronic Health Record)や医療費データ等、日々の医療提供で生成、収集されるデータが中心であり、さらにセルフモニタリングデバイス等から収集されるデータや患者レジストリーデータも含まれる。

なお、FDA RWE のフレームワークは 21st Century Cure Act³⁾ に則ったもので、医薬品の安全性と、限定的範囲での有効性 (ラベリング情報の適応追加・変更、用量や用法、投与経路の変更、等)の評価における RWE の活用が対象とされている。

RWD の活用で、データの品質とデータの標準に焦点が当たっている。RWD のデータソースは診療現場にあることから、RWD/RWE におけるデータ標準の議論は、ヘルスケアと臨床研究・規制領域にわたることが 1 つの重要な特徴である。

3.データ品質

治験におけるデータ品質の議論に比べて、RWD の品質の議論はまだ新しく、多様な議論が展開されている。「データの標準」と「データの品質」はどう関わるのかについては Nahm 等が述べているように、「データの品質」のいかなる点を論じるのかを明確にする必要がある⁴⁾。ここでは RWD/RWE におけるデータ標準を考える上で、データ品質に関する議論をごく限られてはいるが、以下にみてみたい。

3.1 データの品質保証

データの品質保証(Data Quality Assurance)に関しては、広い領域で長年の議論があり、その中でデータの「ベリフィケーション(verification)」と「バリデーション(validation)」という用語がよく使われる。両者の定義は必ずしも一通りではないが、米国環境保護庁(US Environmental Protection Agency)による「環境データ・ベリフィケーションおよびデータ・バリデーションに関するガイダンス」⁵⁾では、次のように定めている。

- データ・ベリフィケーション: データセットの完全性、正確性、適合性(conformance/compliance)を、方法および手続き的要件、契約要求事項に照らして評価するプロセス
- データ・バリデーション: 検査対象(analyte-)およびサンプル固有のプロセスであり、データ・ベリフィケーションを超えて特定のデータセットの分析的クオリティを評価すること

基本的な相違は、データ・ベリフィケーションが分析的手法の手続きや契約要求で事前に定義された(しばしば generic な)要件に照らしての評価であるのに対し、データ・バリデーションはプロジェクト固有の文書における「プロジェクトに特有のデータニーズ」に焦点を当てていることである。一般にデータのベリフィケーションとバリデーションは、シーケンシャルなステップであり、前者が実験室データの収集活動で実施されるのに対し、後者(バリデーション)は、ベリフィケーションのアウトプットから始まる評価であり、ほとんど常にデータ収集者やデータ利用者とは独立のパーティにより実施されるとしている⁵⁾。

FDA RWE のフレームワークではデータの信頼性(reliability)と、データの fitness-for-use, fitness-for-purpose という用語が使われている。データの信頼性(reliability)に関しては、1つのカテゴリが「データ品質管理/品質保証(quality control/quality assurance)とデータ収集プロセス」であり、もう1つのカテゴリが「リサーチ・クエスチョンごとのデータの適切性(relevancy)である」としている。RWD の fitness-for-use は利用に照らしての適合性で、この議論は複雑であり、RWE が規制における意思決定のための fit-for-purpose に必要な1つの要素であるとしている。

データ標準を考える上で、以下ではデータ品質保証のための品質アセスメントについて考える。

3.2 データ品質アセスメント

データ品質アセスメント(Data Quality Assessment: DQA)に関しては研究、組織・団体等による開発等^{6), 7)}がなされており、ターミノロジーも多様である。Kahn 等⁸⁾はバイオメディカル・インフォマティクスにおける既存のターミノロジーを単一化させる(unify)ことで、調和された DQA ターミノロジーを整理・提案している。そこでは以下の3つのカテゴリを示している。

- a) Conformance(適合性): データが以下に適合する度合い。
 - ①あらかじめ定義された内的または外的なフォーマット・制約への適合 (Value Conformance)、②構造的な制約への適合。例えば主キーとフォーリンキーの関係等、③既存の値に基づいて計算により導出された値の正確性 (Calculation Conformance)。
- b) Completeness(完全性): データの存在が①時間軸の一点 (Atemporal Completeness)と、②複数の時点 (Temporal Completeness)の二つの観点からアセスされること。
- c) Plausibility(データの質が合理的で真であるらしい度合い the quality of being reasonable and likely to be true): データが信じるに足る程度が次のようにアセスされること。①論理的に制約された測定や分布、同一の事実についての独立な複数の測定が一致すること (Atemporal Plausibility)、②時間軸の属性や順序および測定値の状態遷移 (Temporal Plausibility)、③重複する測定値、変数の値あるいは記録の存在 (Uniqueness Plausibility)。

さらに8つのサブカテゴリが定義されている。調和された DQA ターミノロジーは、データに内在する DQA カテゴリ (データの値そのものに属する次元)に限定しており、外来性(extrinsic)のカテゴリ、すなわちデータが利用されるコンテキストに依存する次元、fitness-for-use 基準のデータの利用可能性やタイムリネス、等の役割は含んでいない。データ品質の議論は用語自体、単純ではないが、調和された DQA ターミノロジーに準じることでデータ標準の議論も整理し易いと考えられる。

データ・バリデーションのコンテキストからのデータ品質チェックは重要であるが、データ・ベリフィケーションのコンテキストにおけるチェックに比べて、定義、実施とも複雑である。適合

性 (Conformance) はデータの相互利用可能性 (Interoperability)に強く関連している。ここでは、データ・ベリフィケーションのコンテキストにおけるデータ品質の観点から、データの Interoperability を中心にデータ標準について考える。なお、RWD の完全性(Completeness)は良く知られているように、RCT で前向きに収集されたデータに比べて低い。

4. データ Interoperability をコアとするデータ標準の取り組み

近年 HL7 FHIR(Fast Healthcare Interoperability Resource)が急速に展開している。臨床試験、規制領域においては CDISC における HL7 FHIR の取り組みや、米国医薬品食品局(FDA)、欧州医薬品庁(EMA)から参加しての医薬品あるいは臨床研究の FHIR リソース開発など顕著な活動がある。臨床研究への RWD 活用の一環として、コモンデータモデル (Common Data Model: CDM)の取組みがある。

4.1 コモンデータモデル

4.1.1 研究ネットワークとコモンデータモデル

以下をはじめとする様々な研究ネットワークが展開する中、それぞれのコモンデータモデルが策定されている⁹⁻¹³⁾。

- 1) PCORNet(national patient-centered clinical research network) コモンデータモデル
- 2) i2b2 (Informatics for Integration Biology and Bedside) コモンデータモデル
- 3) OMOP(Observational Medical Outcomes Partnership) コモンデータモデル
- 4) Sentinel model

以下、1)と2)について簡単に述べる。PCORNet(全国患者中心臨床研究ネットワーク)⁹⁾は、2014年、患者中心アウトカム研究所 (Patient Centered Outcomes Research Institute: PCORI) が2億5千万ドル以上を投資して開発された。PCORNet は迅速・容易、低コストでニーズに適う臨床研究が実施できるように設計された。「国全体に渡る代表性の高いパートナー・ネットワークからなるネットワーク」で、パートナー・ネットワークには臨床データ研究ネットワーク (Clinical Data Research Networks: CDRNs)と「患者が強化する研究ネットワーク (Patient Powered Research Network: PPRN)」が含まれる。臨床研究を研究者指向型から患者中心型の研究へとシフトさせていくもので、患者、ケア提供者、その他のステークホルダーが研究のデザイン、実施、分析から結果の普及に至るまで関与することを支援している。

Informatics for Integration Biology and Bedside (i2b2)は、NIH のファンドによる Biomedical Computing のためのナショナルセンターであり 2004 に設立された。ボストンの Partners HealthCare System に本拠がある¹⁰⁾。i2b2 にはハーバード連携病院、ハーバード公衆衛生大学院、ハーバード医学大学院、ハーバード/MIT ヘルスサイエンス&テクノロジー部門の研究者が参加している。国のサイエンス・コミュニティへの i2b2 ツールの提供がなされ、さらなる推進が期待されている。

4.1.2 Common Data Models Harmonization (CDMH)

様々な CDM が存在する中で、「エビデンス創出のための各種 CDM とオープンスタンダードのハーモナイゼーション」とよばれる取り組みが連邦機関の主導によりなされている¹⁴⁾。日々の医療提供により生成される観察的データを用いて、患者中心のアウトカムリサーチを行うためのデータ基盤を構築するもので、連邦機関として、FDA の他に米国国立衛生研究

所の国立医学図書館(NLM)、国立がん研究所(NCI)、国立先端トランスレーショナル科学センター(NCATS)、米国国家医療 IT 調整官室(Office of the National Coordinator for Health Information Technology: ONC)が参加している。

コモンデータモデル・ハーモナイゼーション(Common Data Models Harmonization: CDMH) FHIR IG は、HL7 の BR&R (Biomedical Research and Regulation) Work Group により開発されている。現在の版(version 0.1.0)は、まだ承認を得たものではなく継続的に開発されている。複数の研究ネットワークから抽出される研究データは、それぞれが異なるデータモデルを用いてデータ表現され得ることから、本プロジェクトは、PCORnet CDM、i2b2 CMD、OMOP CDM、センチネルモデル(Sentinel model)に焦点を当てている。同 FHIR IG は患者中心のアウトカムリサーチ(PCOR)から抽出された観察データの FHIR フォーマットへのマッピングと変換に焦点を当てている。

連邦機関と研究ネットワークによる CDMH プロジェクトでは、データクエリーや、クエリーから結果を得るなど、様々な観点から取り組んでいるが、CDMH FHIR IG は、クエリーから得られる結果の FHIR フォーマットへのマッピングと変換だけに焦点を当てている。Learning Health System (LHS)や PCOR Trust Fund (PCORTF) による DAF-Research IG (Data Access Framework: DAF)、構造化データキャプチャ(Structured Data Capture: SDC) IG、患者報告アウトカム FHIR IG、Women's Health Technologies Coordinated Registry Network (CRN) FHIR IG などへの投資とともに、米国のデータ基盤の向上を目指している。CDMH FHIR IG は将来、クエリーやワークフローが自動化可能となる段階において、DAF-Research IG を向上させることが期待される。

4.2 規制当局の主導による推進

詳細は省略するが、FDA と EMA は product and substance identification のための FHIR implementation の開発を連携して行っている。FHIR リソースでは臨床研究のための ResearchStudy、ResearchSubject などの開発が進められている。FDA は、Harvard Pilgrim Health Care Institute とハーバード医学大学院と協力して、複数データソースにわたる EHR データの質、完全性、安定性を評価し、記述するための統一されたメタデータ標準のため組んでいる。複数の連邦機関、研究者が患者中心のアウトカムリサーチのために RWD を適切に活用できるようにするため、複数のデータモデルのハーモナイゼーションのためのメタ・コモンデータモデルを開発している。RWD/RWE の促進のため、規制当局が主導する形で、データ標準のハーモナイゼーションへの取り組みがなされている。

5.おわりに

RWD に関しては、複数施設からデータを収集して行う RWD による観察研究では、どのようなデータであるのか、データの品質をどう保証するのかが問われる。複数施設が参加する研究ネットワークで、コモンデータモデルが定義される。コモンデータモデルは、データの適合性(compliance)の評価に必要であり、データの interoperability の基本となる。RWD の研究活用では、一つの研究ネットワーク内だけにとどまらず、複数ネットワークにわたる活用へと広がっていく。そこでは複数のコモンデータモデルのハーモナイゼーションの取り組みがなされている。単一組織から複数組織のネットワークへ、ネットワークとネットワークを繋ぐ、さらなるネットワークへと展開する中で、データモデルからコモンデータモデル(CDM)へ、

さらにコモンデータモデルのハーモナイゼーション(CDMH)への展開がある。RWD 活用のための標準の動向として、臨床研究、医薬品規制領域においても顕著な FHIR の活動がある。RWD の追求において医療情報に求められるのは品質に支えられた、相互利用可能性を保証する interoperable なデータであると言える。精緻な議論の積み重ねと、多大な努力・連携・協調・調和の上に、世界は RWD/RWE の本格活用の時代に向かっている。

謝辞 本研究は、AMED の課題番号 JP20mk0101103 の支援を受けた。

参考文献

- 1) The International Council for Harmonization. ICH Reflection on “GCP Renovation”: Modernization of ICH E8 and Subsequent Renovation of ICH E6, Jan 2017. [https://admin.ich.org/sites/default/files/inline-files/ICH_Reflection_paper_GCP_Renovation_Jan_2017_Final.pdf] (cited 2020-Jun-1)]
- 2) U.S. Food and Drug Administration. Framework for FDA's Real-World Evidence Program. December 2018. [<https://www.fda.gov/media/120060/download>] (cited 2020-Jun-1)]
- 3) H.R.34 - 21st Century Cures Act. [<https://www.congress.gov/bill/114th-congress/house-bill/34>] (cited 2020-Jun-1)]
- 4) Nahm M, Hammond WE. Data Standards ≠ Data Quality Stud Health Technol Inform. 2013; 192: 1208.
- 5) United States Environmental Protection Agency. Guidance on Environmental Data Verification and Data Validation, EPA QA/G-8. [<https://www.epa.gov/sites/production/files/2015-06/documents/g8-final.pdf>] (cited 2020-Jul-1)]
- 6) Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc JAMIA. 2013;20(1):144-51.
- 7) Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. A Data Quality Ontology for the Secondary Use of EHR Data. Proc AMIA Symp. 2015:1937-46.
- 8) Kahn M, Callahan T, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016;4(1):1244. doi: 10.13063/2327-9214.1244. eCollection 2016. [<http://repository.edm-forum.org/egems/vol4/iss1/18>] (cited 2020-Jun-1)].
- 9) Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc JAMIA. 2014;21(4):578-82.
- 10) Klann JG, Joss MAH, Embree K, Murphy SN. Data model harmonization for the All Of Us Research Program: Transforming i2b2 data into the OMOP common data model. PLoS One. 2019; 14(2):e0212463. Published online 2019 Feb 19. [<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6380544/>] (cited 2020-Jun-1)]
- 11) Assessments of Health Outcomes Among Individuals Exposed to Medical Products | Mini-Sentinel. [http://mini-sentinel.org/assessments/medical_events/default.aspx] (cited 2020-Jun-1)]
- 12) Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System - A National Resource for Evidence Development. N Engl J Med. 2011; 364(6):498-9.
- 13) Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data

Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015; 216:574-8.

- 14) Office of the Assistant Secretary for Planning and Evaluation. Harmonization of various common data models and open standards for evidence generation. [<https://aspe.hhs.gov/harmonization-various-common-data-models-and-open-standards-evidence-generation> (cited 2020-Jul-30)]