RWD140

Benefits and Challenges of Deterministic, Probabilistic, and Patient-Mediated Approaches to Linking Patients' Clinical Trial Data (CTD) and Their Real World Data (RWD)

Zabell T[1], Jeswani N[2], Hahn K[3], Coutcher J[3], Papsch R[1]

[1]IQVIA, London, LON, UK; [2]RxPx, London, UK; [3]IQVIA, Cambridge, MA, US

OBJECTIVES

- Whilst the clinical depth and internal validity of Clinical Trial Data (CTD) is unparalleled, Real World Data (RWD) can expand our view multi-dimensionally, over time and across broader patient settings (Figure 1).
- CTD and RWD are already widely considered to be mutually complementary, however, the ability to combine both forms of health data together, at a patient-level, presents the opportunity to gain insights that help to build a richer understanding of drug value[1].
- Linkage of patients' CTD and RWD can provide longitudinal insights into their disease and treatment journey and associated outcomes before, during, and after clinical trials. These linked data can span the settings of primary care, pharmacies, emergency care, specialist tertiary centers, and in the patient's home or long-term care facility[2].
- We aimed to understand the different methodologies for linking CTD and RWD, characterising use cases, strengths, limitations, and operational considerations with each approach.

METHODS

- Through pragmatic review of the literature and ongoing in-house research, we identified precedent of CTD-RWD linkages, and RWD-RWD linkages with an eye toward models adaptable to CTD.

RESULTS

- We identified three predominant approaches for linking CTD and RWD. The suitability of these approaches is dependent on many interacting factors such as the intended use case for data linkage, the applicable information governance (IG) environment, and the relevant RWD landscape.
- Within the RWD landscape, the capture and completeness of variables available to link, and the extent to which they are individually and collectively able to uniquely identify an individual patient will be key success factors for a feasible linkage effort[3].
- From an operational perspective, success hinges on aligning the design principles to the anticipated evidence needs through timely and strategic planning of the research governance, data access and consent models, directionality of patient recruitment, and technology-enablement.

1. Deterministic

- In deterministic linkage (Figure 2), CTD and RWD are directly connected by unique patient identifier fields, such as national insurance or social security numbers.
- Since the patient identifiers used are unique, the deterministic approach should always produce a 1:1 match where a patient exists in both the CTD and RWD, thus the success rate for linkage should be dependent only on the level of patient overlap between the clinical trial and real world populations captured.
- This linkage approach is suitable in cases where the use of identifiable patient data is acceptable within the IG framework. A case study of deterministic linkage used patients' Medicare Beneficiary Identifier (MBI) to link their CTD to Medicare claims data (RWD) to study healthcare resource use and costs of care among Medicare-enrolled patients during and after participating in clinical trials for early Alzheimer's disease[4]. The site-based recruitment approach achieved a 75% acceptance rate - prioritising institutional review board approval of the study protocol and informed consent form before site enrolment assured sites about the privacy measures in place. Nevertheless, sites who declined mainly cited concerns over requesting that patients allow access to their Medicare data, a concern that contrasts with a previous study conducted in patients with cancer, in which 93% of those surveyed were willing to allow long-term access to their claims data[5].
- The main challenge and consideration with a deterministic approach is protecting patient privacy; identifiable patient-level RWD is often not able to be shared, particularly in Europe.

2. Probabilistic

- Tokenization involves the encryption of personally identifiable information (PII) into "tokens" to pseudonymize patient-level RWD fields such as name, date of birth, and healthcare provider postcode. The tokenized RWD can then support probabilistic matching (Figure 3) with a clinical trial dataset that has undergone an analogous tokenization process. The probabilistic matching uses a statistical approach to measure the probability that two patient records represent the same individual.
- Success rates of probabilistic linkage are typically expected to be lower than deterministic as the linkage is dependent not only on the patient population overlap between CTD and RWD, but also the matching methodology applied, and the unique identifiability of the data used to generate the tokens.
- Many variations in the methodology exist for probabilistic matching, which can be a technically complex and computationally intensive process. The underlying aim is to minimise false positives and false negatives. The probabilistic linkage threshold (the acceptable likelihood that two records are a true match based on whether they agree or disagree on the various identifiers) should be chosen as such, with the appropriate trade-offs between sensitivity and specificity. Unsatisfactory matches should be adjudicated by reference to additional information.
- Other best-practices include building multiple layers of hashing into the tokenization process to increase the number of steps and protect against decryption, controlling for missing or out-of-date data, and ensuring additional privacy safeguards are in place to mitigate the increased risk of re-identification presented by linkage.
- Successful probabilistic linkage of CTD-RWD has been demonstrated in the US, with wide-scale efforts to tokenize and de-identify PII in RWD underway, and the approach is emerging in Europe and Canada. A case study of probabilistic linkage demonstrated the feasibility of probabilistically linking retrospective medical claims and longitudinal prescription claims data to CTD to increase the depth and breadth of research capabilities in rheumatoid arthritis[6]. Patients were matched on age, gender, and the three-digit ZIP code of their provider, applying a point scoring system using additional clinical variables such as National Provider Identifier, use of disease-modifying anti-rheumatic drugs, hospital visits and diagnosis dates. Of the 78 patients consented, 44 (56%) were successfully matched (met the probabilistic matching threshold) across the 3 data sources.

3. Patient-mediated

- In direct patient-mediated access to RWD (Figure 4), patients consent for third-party collection and integration of their RWD with their CTD. The consent process is adapted for this additional consent (and crucially should not be burdened), and the RWD accessed and linked enters the same data management and trial reporting pathways.
- The differentiating factor and main benefit of patient-mediated access over more traditional pathways of acquiring RWD is that access to patient data is being governed by patients themselves rather than providers. This more direct involvement of patients in the governance of their data, gives them a clearer opportunity to understand how and why their data will be used, so they can make a more informed decision on whether or not they are comfortable with the use of their identified data. Further, coverage is typically going to be higher in this model, since the trial population and the linked population are identical. Patient-centricity lies at the core of the patient-mediated approach, so best practice is to share insights back with patients to help them understand their disease and see the benefits of sharing their data.
- The situation for patient-mediated data access varies significantly across geographies. The US is furthest ahead internationally, with provider obligations and incentives via tweaking Medicare reimbursement regulations to share patient-requested data under The 21st Century Cures Act[7], leveraging FHIR (Fast Healthcare Interoperability Resources)-based interoperability. Other countries, including the UK, France, and the Nordic region, are in the process adopting similar approaches, with precedent of both patient-mediated access to national RWD and linkage between datasets. The EU is also working towards a common file format to facilitate interoperability.

CONCLUSIONS

Realising the benefits of CTD-RWD linkage will involve learning from experience and maturation of the field, however, as a starting point, there are fundamental principles that can guide value-adding research from the outset. All required consent should be sought at trial enrolment and should not delay trial operations. The research objective and use case should be aligned with the linked cohort sizes achievable, to optimise the use of time and resources. Crucially, the burden of CTD-RWD linkage on providers and patients should be minimised. Technology-enabled solutions should be leveraged to overcome the challenges of linkage whilst protecting data privacy. As a top-down solution design, standardised tokenisation engines on RWD can enable scalable linkage whilst privacy analytics help mitigate re-identification risk. As a bottom-up solution design, patient-mediated access empowers patients to have autonomy over their data in the increasingly digitalised world of healthcare and could exponentiate the value of insights by connecting multi-dimensional data for each individual patient.

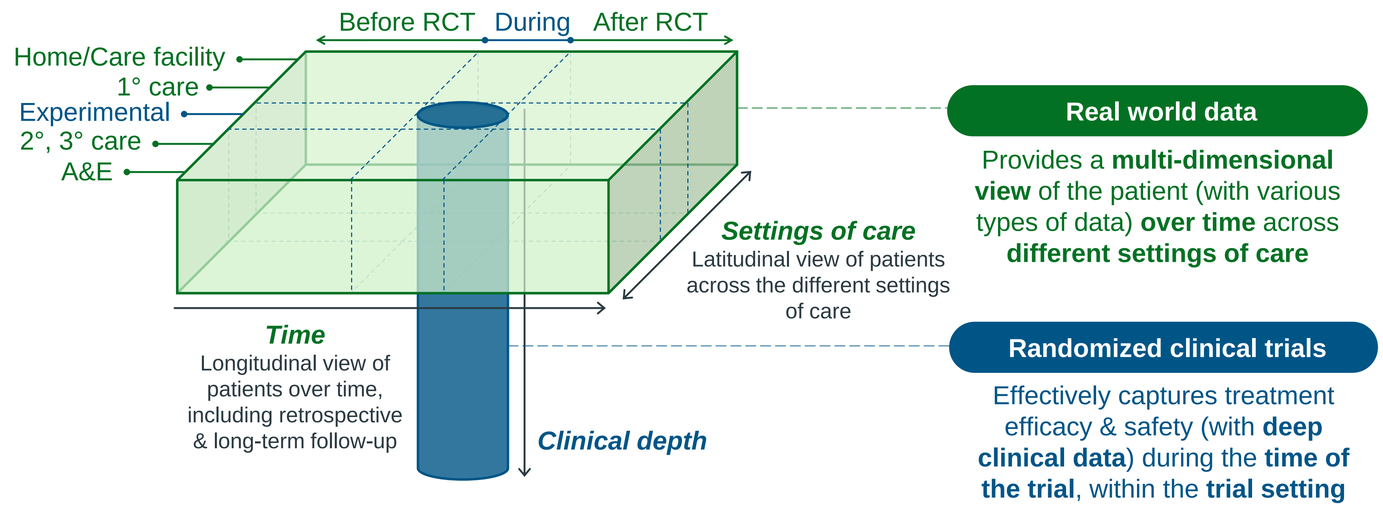

Figure 1 – Comparative coverage of real world data vs clinical trial data for the dimensions of time, settings of care, and clinical depth

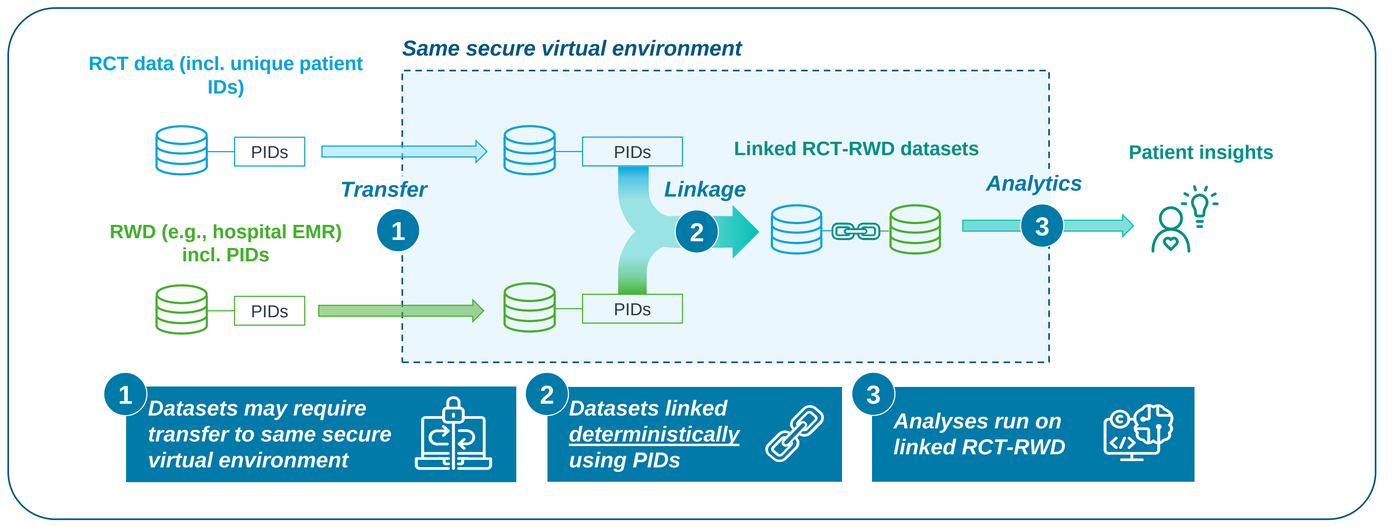

Figure 2 – Illustrative example of deterministic linkage to connect clinical trial data and real world data via patient identifiers

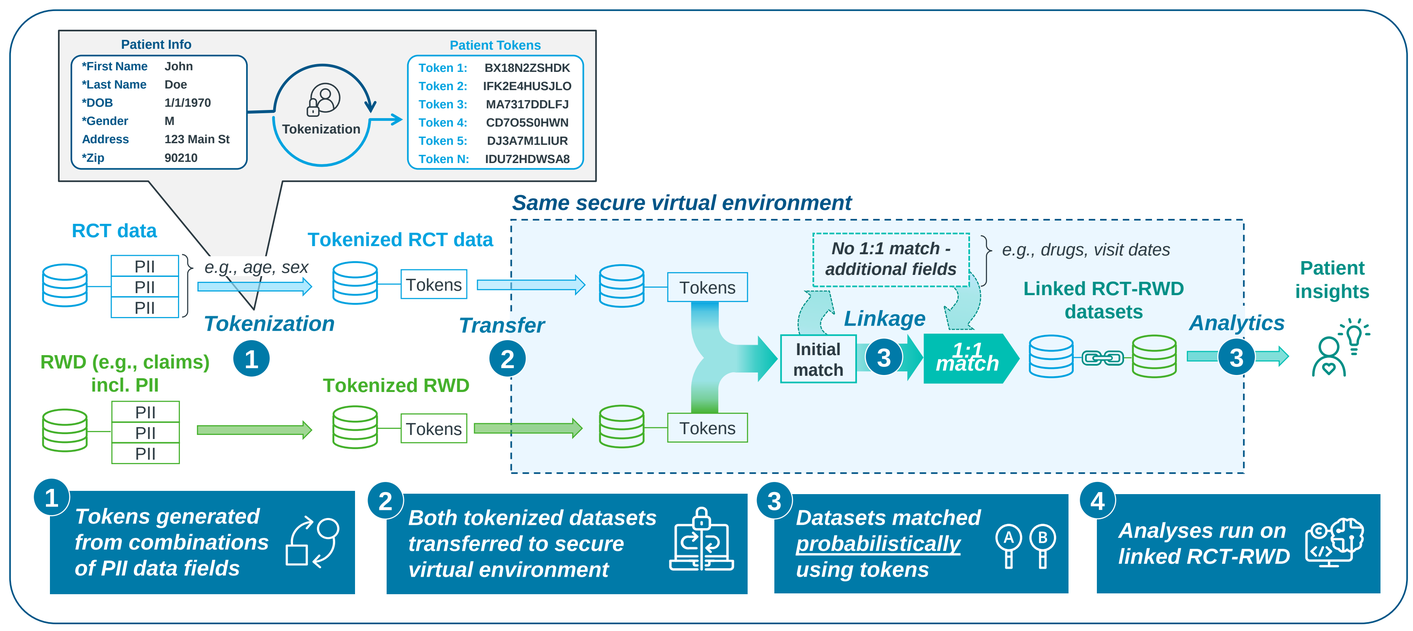

Figure 3 – Illustrative example of how tokenization can be used to support probabilistic matching

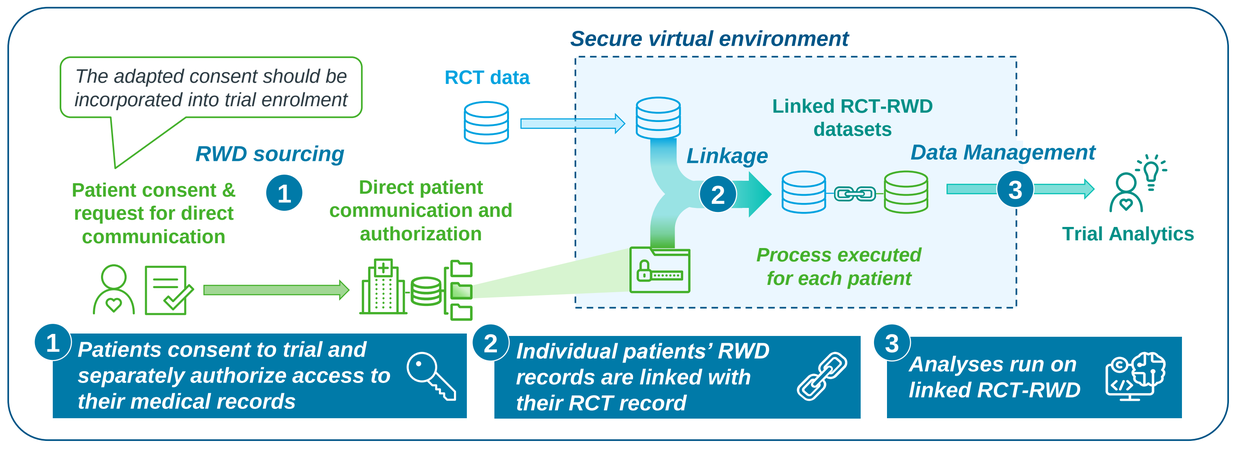

Figure 4 – Illustrative example of additional patient consent workflow for direct patient-mediated access to RWD for integration with clinical trial data

1. Kim, H. S., Lee, S., & Kim, J. H. (2018). Real-world Evidence versus Randomized Controlled Trial: Clinical Research Based on Electronic Medical Records. Journal of Korean medical science, 33(34), e213. https://doi.org/10.3346/jkms.2018.33.e213
2. FDA (2020). Enhancing evidence generation by linking randomized clinical trials (RCTs) to real world data (RWD). https://reporter.nih.gov/project-details/10185038
3. Dusetzina SB, Tyree S, Meyer AM, et al. Linking Data for Health Services Research: A Framework and Instructional Guide. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Sep.
4. Abbass et al. (2020). A Novel Approach to Study Healthcare Resource Use Costs of Care Among Medicare-Enrolled Patients During and After Participating in Clinical Trials For Early Alzheimer's Disease. *Alzheimer's & Dementia: The Journal of Alzheimer's Association.* **15** (7), P1574-P1575. aanddjournal.net/article/S1552-5260(19)35379-8/pdf
5. Cheema PK, Rothenstein J, Melosky B, Brade A, Hirsh V. Perspectives on treatment advances for stage III locally advanced unresectable non-small-cell lung cancer. Curr Oncol. 2019 Feb;26(1):37-42. doi: 10.3747/co.25.4096. Epub 2019 Feb 1. PMID: 30853796; PMCID: PMC6380636.
6. McGuiness et al. (2021). Probabilistic Linkage of Randomized Controlled Trial Data to Administrative Claims: A Case Study of Patients from Baricitinib Clinical Trials. Rheumatology and Therapy. Jun; 8(2): 793–802. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8217382/
7. Lye, C. T., Forman, H. P., Daniel, J. G., & Krumholz, H. M. (2018). The 21st Century Cures Act and electronic health records one year later: will patients see the benefits?. Journal of the American Medical Informatics Association : JAMIA, 25(9), 1218–1220. https://doi.org/10.1093/jamia/ocy065

ISPOR Europe 2022, Nov 6-9, Vienna, Austria