Best Practices for Developing Disease-Specific Federated Networks: Insights from a Systemic Lupus Erythematosus Study

Clair Blacketer[1,2], Frank DeFalco[1], Gowtham A Rao[1], Anna Sheahan[1], Michel Van Speybroeck[1], Martine Lewi[3], Federico Zazzetti[4]

[1]Janssen Research & Development, Raritan, NJ, 2 Department of Medical Informatics, Eramus MC, Rotterdam, NL, 3 Janssen Global Services, LLC, Raritan, NJ, USA, 4 Janssen Global Services, LLC, a Johnson & Johnson Company, Immunology Medical Affairs, Horsham, PA, USA

Background

The development of disease-specific federated networks (FN) poses unique challenges and opportunities in the realm of data standardization, quality assurance, and collaborative analysis[1]. We explore best practices for establishing these networks and how they can be utilized to generate evidence, focusing on the experiences and insights gained from a federated network of five global data sources developed to study Systemic Lupus Erythematosus (SLE) through standardization to the OMOP Common Data Model (CDM)[2]. The critical aspects covered include data quality and characterization, leveraging standardized vocabularies, database heterogeneity, and collaborative engagement with network partners.

Methods

Standardization of Data

Data standardization, both structural and semantic, plays a crucial role in the efficiency and effectiveness of federated networks. Structurally, utilizing uniform tables and fields of the CDM across registries allows for the development of queries and methods that can be executed uniformly. This uniformity reduces the need for de novo query creation, which can be labor-intensive and prone to inconsistencies.

Semantically, the use of standard vocabularies offers significant advantages. In the context of the SLE study, the process of defining clinical characteristics of patients at registry enrollment was initially hindered by the use of differing methods created to represent the same questionnaires across the data sources. For example, four of the registries include the Systemic Lupus Erythematosus Daily Activity Index (SLEDAI). As shown in table 1, the question on hematuria was not mapped to a standard concept in registry 1, it was mapped similarly in registries 2 and 3, and in registry 4 only the answer ‘no’ was mapped. Upon identifying these issues all SLEDAI mappings were aligned.

To streamline evidence generation and promote interoperability, federated network participants should adhere to the same vocabulary rules when mapping the same or similar clinical events.

Table 1: Questionnaire mapping example across four registries prior to alignment on vocabulary rules

Registry 1	Registry 2	Registry 3	Registry 4	Target Concept	Target Domain	Value As Concept
	SLEDAI-2K Descriptor | Haematuria | >5 red blood cells/ high power field | No	SLEDAI activity index: Hematuria | No		2000004167	Measurement	0
		SLEDAI activity index: Hematuria | Unknown		2000004166	Measurement	0
	SLEDAI-2K Descriptor | Haematuria | >5 red blood cells/ high power field | Yes	SLEDAI activity index: Hematuria | Yes		2000004168	Measurement	0
	SLEDAI-2K Descriptor | Haematuria | >5 red blood cells/ high power field | Yes	SLEDAI activity index: Hematuria | Yes		1340204	Observation	79864
	SLEDAI-2K Descriptor | Haematuria | >5 red blood cells/ high power field | No	SLEDAI activity index: Hematuria | No	Hematuria (SLEDAI) | no	40481925	Observation	79864
		SLEDAI activity index: Hematuria | Unknown		4287024	Observation	79864
SLEDAI-2K | Hematuria | >5 red blood cells/ high power field				-	-	-
			Hematuria (SLEDAI) | yes	-	-	-

Data Quality and Characterization

Effective data quality assurance and characterization are paramount when establishing a federated network. During the early stages of the SLE network, it was observed that registry partners struggled to interact meaningfully with the data once it was standardized to the CDM, leading to challenges in identifying and addressing data quality concerns. To mitigate these issues, it is essential to conduct thorough and explicit data quality investigations. These investigations should be presented in a format that is easily comprehensible to all stakeholders, considering the diverse nature of the native data sources involved.

Regular readouts and structured feedback sessions with participants can significantly enhance their understanding and ability to identify data quality issues. Such proactive

measures ensure that data cleaning and quality control processes are streamlined and more effective, thereby facilitating smoother subsequent analyses.

Generating Evidence Using Standardized Registry Data

Generating evidence using standardized registry data is challenging as it requires addressing the inherent variability and complexity of the data sources. In federated networks, individual data partners often collect data through different instruments, each with its own semantic nuances and levels of granularity. These variations can pose significant challenges for data integration and analysis.

For instance, differences in granularity between registries may result in varying levels of detail for similar data points. This is shown clearly in table 1. Even though all four registries collect the SLEDAI, only one allows the answer ‘unknown’. Additionally, temporal discrepancies can arise when data is collected at different intervals or points in time across registries. Furthermore, underlying heterogeneity in the inclusion and exclusion criteria of the registries complicates data harmonization. Each registry may have multiple sets of criteria that are not easily captured in the CDM structure as many do not have dates associated with the conditions used to identify and recruit patients.

To effectively generate evidence, it is crucial to acknowledge and address these differences. This can involve comprehensive database characterization to facilitate fitness-for-use evaluations, developing techniques to represent the inclusion and exclusion criteria for each registry in the OMOP CDM, and critically assessing common data elements that can be leveraged in network-based analyses. By doing so, researchers can better understand the context and limitations of the integrated data, leading to more accurate and meaningful analyses.

Collaborative Engagement

Engaging network partners in a collaborative and inclusive manner is crucial for the success of federated networks. A key learning from the SLE study was the importance of helping network partners understand the definitions and methodologies behind analyses, given their familiarity with their own data. For instance, aligning the mappings of clinical concepts across different registries involved in the SLE study required a detailed examination of the source-to-concept mapping information.

By consolidating this information, discrepancies such as the inclusion of 'unknown' response options in some registries but not others were identified and addressed. Ensuring that all registry partners are on the same page with regard to data mappings and methodologies prevents inconsistencies and enhances the reliability of the overall analysis.

Recommendations for Future Studies

Based on the experiences and insights from the SLE federated network study, several recommendations can be made for future efforts:

1. **Enhanced Data Quality Investigations**: Conduct comprehensive data quality checks and present findings in an accessible format for all stakeholders.
2. **Leverage Standard Vocabularies**: Utilize common concepts and mapping techniques to align the data sources and promote easier evidence generation.
3. **Address Data Variability**: Develop techniques to manage differences in data granularity, temporality, and registry inclusion/exclusion criteria.
4. **Foster Collaborative Engagement**: Engage registry partners in ongoing discussions to ensure a common understanding of data mappings and analytical definitions.
5. **Iterative Improvements**: Continuously refine data standardization efforts and methodologies based on feedback and evolving best practices.

By adhering to these best practices, disease-specific federated networks can achieve more reliable, accurate, and meaningful outcomes, ultimately advancing the understanding and treatment of various conditions.

References

1 van Baalen V, Didden E-M, Rosenberg D, *et al.* Increase transparency and reproducibility of real-world evidence in rare diseases through disease-specific Federated Data Networks. *Pharmacoepidemiol Drug Saf*. 2024;33:e5778. doi: 10.1002/pds.5778

2 Schreiber J, Speybroeck MV, Zazzetti F, *et al.* Pos1501 Lupusnet – a Federated Model/Network to Support Real-World Data Research in Systemic Lupus Erythematosus. *Annals of the Rheumatic Diseases*. 2023;82:1108–1108. doi: 10.1136/annrheumdis-2023-eular.3017