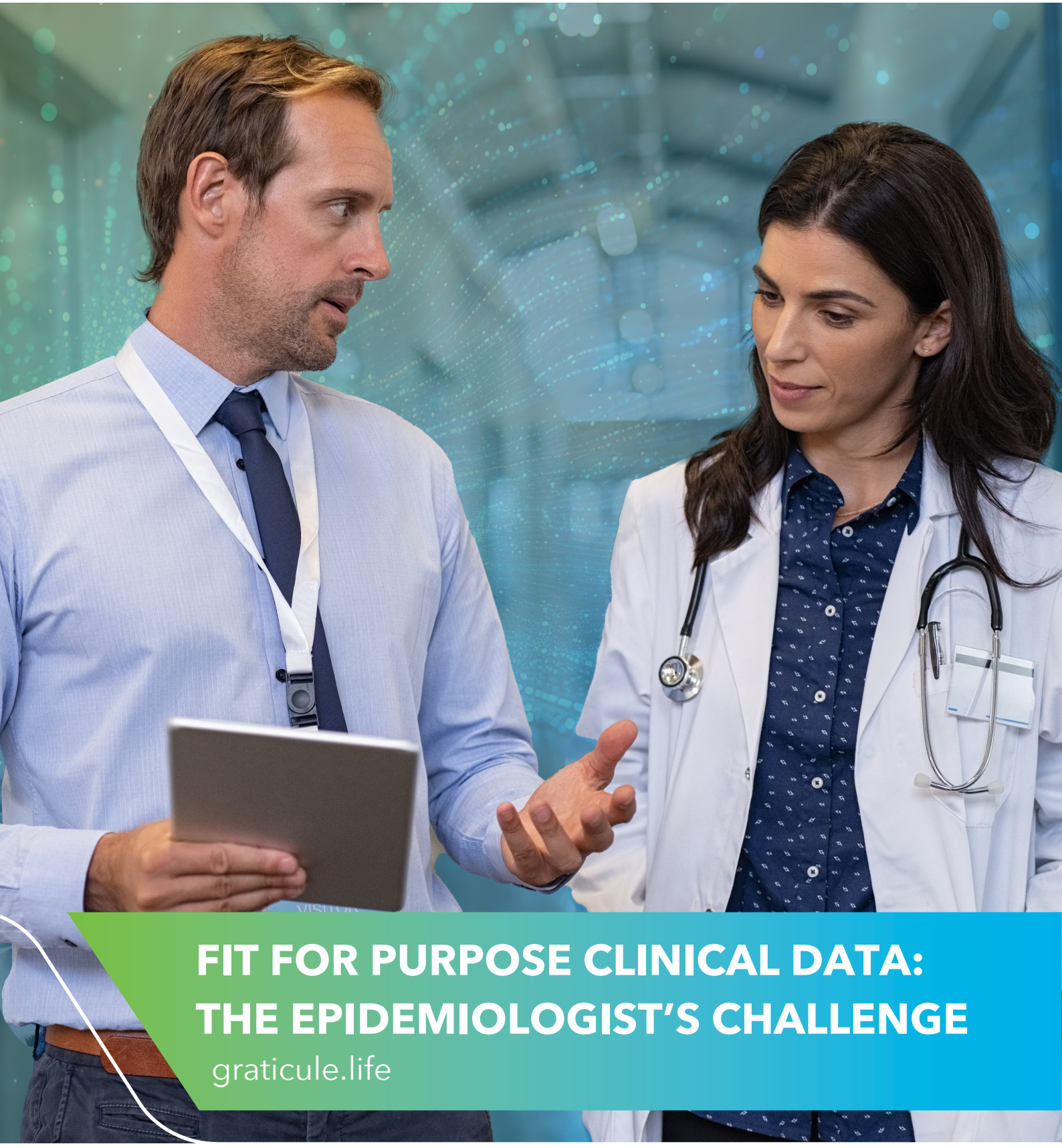
graticule™

White Paper



FIT FOR PURPOSE CLINICAL DATA: THE EPIDEMIOLOGIST'S CHALLENGE

graticule.life

Overview

For life sciences organizations, the secondary use of patient data plays a decisive role in real world evidence generation.

Ensuring that healthcare decisions are informed by accurate, relevant, and timely patient data is essential, from the development of effective treatments and public health initiatives to commercial market access and drug reimbursement strategies. Navigating the complex landscape of accessing local, country-level patient data in Europe has become increasingly critical. However, obtaining this data at scale and in a cost-effective manner is fraught with challenges, as aggregate off-the-shelf data sources that are commonplace in the United States (e.g., claims, electronic health record) are almost nonexistent within the regulatory framework of the European Union (EU). In turn, Real World Data (RWD) in Europe is not as commoditized as in the US and is almost always owned by public institutions rather than the private sector. Key differences within EU countries such as regulations, clinical practice, medical coding, culture and language lead to a variable RWD landscape. Thus, life science organizations must seek alternate pathways to data access.

Introduction:

While most organizations recognize the importance of leveraging local European patient data, few are fully equipped to overcome the significant barriers to data access on a large, multi-purpose, and cost-effective scale. This white paper discusses the current state of secondary data access in Europe and explores key barriers to access faced by life science organizations, including those due to complexities of regulatory compliance, heterogeneous care delivery, and variability in data quality and standards. Finally, the discussion includes strategies to overcome these barriers, enabling organizations to harness the full potential of such data at scale to advance healthcare across Europe.

Importance of Accessing Local Patient Data in Europe:

Variability in care delivery, public policies, and even geography can have profound impacts on care outcomes. For example, the generalization of regional health data to a different population may lead to inaccurate conclusions when applied to the new population. Increasingly, regulatory and HTA bodies in the EU prefer high quality, locally relevant data not only to understand the burden of disease and the patient journey, but also to evaluate safety, efficacy, and cost-effectiveness of new medical treatments, all of which may vary significantly across regions. HTA bodies, which play a critical role in determining the reimbursement and pricing of new therapies, particularly rely on local data to assess the value of new treatments within the specific healthcare context of each country. HTA assessments often require real-world evidence from the local population to make informed decisions about the cost-effectiveness and budget impact of new interventions. While practices vary, HTA bodies in France (HAS) and Germany (IQWi G, G-BA) rarely accept non-local data on their own, whereas the UK (NICE), Italy (AIFA) and Spain (AEMPS) are typically less stringent. Nordics generally accept data from other Nordic countries, and smaller EU countries (e.g., Netherlands) are more amenable to non-local data but are increasingly requiring at least some locally representative data.

References:

PharmacoEconomics - Open, "The Acceptance of Indirect Treatment Comparison Methods in Oncology by Health Technology Assessment Agencies in England, France, Germany, Italy, and Spain." Available from: SpringerLink.

Health Economics Review, "Relevance of indirect comparisons in the German early benefit assessment and in comparison to HTA processes in England, France and Scotland." Available from: BioMed Central.

Optimax Access, "HTA in Five EU Countries: UK, France, Germany, Spain, and Italy." Available from: Optimax Access.)

This is underscored by the 2021 adoption of the EU HTA regulation (HTAR), set to be fully implemented by 2025, which aims to streamline and harmonize the health technology assessment process across EU member states – emphasizing the need for locally relevant data to inform joint clinical assessments (JCAs). The HTAR is particularly focused on ensuring that the data used in these assessments reflects the diverse populations across Europe, considering factors such as local disease burden and variations in patient care pathways. This approach helps avoid the pitfalls of generalized data that might not be applicable in specific national contexts.

Current Routes to Data Access

With strategic and financial considerations in the balance, Life Science Organizations have traditionally relied on two routes to data access, with a third quickly gaining traction:

Data Registries:

One of the most straightforward methods of accessing patient data in Europe is through public data registries. These registries are centralized databases that systematically collect, store, and manage patient data across specific populations or health conditions. The Nordic countries (Denmark, Sweden, Norway, and Finland) are particularly renowned for their comprehensive and high-quality health data registries. These registries often contain longitudinal data on large cohorts, providing invaluable insights into disease trends, treatment outcomes, and population health. For example, the Danish National Patient Register and the Swedish National Patient Register are widely used by researchers and healthcare organizations due to their extensive coverage and detailed patient records. These registries allow for robust epidemiological studies and can be a goldmine for understanding long-term health outcomes. Registry adoption of best-practice data models and maintenance standards also contribute to study efficiency. Importantly, these two registries represent just a small component of the total set of registries- each Nordic citizen is furnished with a unique PIN, enabling researchers to link patients across datasets/registers for deep high-quality longitudinal coverage. However, the availability of such registries is limited to a few countries – something that is particularly troubling for questions of regional patient journeys and burden of disease. Data access to these registries is often slow (~ 6 to 12 months), and in many cases data feasibility cannot be carried out until access has been approved. Outside the Nordics, similar comprehensive data registries are not as commonplace and linking registries is more difficult, making this route less accessible for those seeking data across broader European regions. Additionally, while registries offer valuable insights, they may lack the granularity or specific data points needed for certain types of research - leading to a 'lowest common denominator' dataset.

Individual site access:

Direct site access may be a necessary approach when study data is not accessible through registries—either altogether or within the requisite timeframe. This method involves negotiating access to data from hospitals, clinics, or other healthcare providers, often for a specific research study or to meet regulatory requirements. This approach allows for the collection of highly specific data tailored to the needs of the study. However, this method also comes with significant challenges, many of which stem from working with public sector entities. Unlike commercially-available data sources in the U.S., public sector institutions often operate under different timelines, incentives, and priorities. This disparity can complicate access to data, protracting the process out of study scope. Generally, while sites are often well versed in prospective research, secondary patient data use for observational studies may be less familiar to IT personnel and ethics committees, leading to strict interpretations of GDPR and other regulatory guidelines that ultimately compromise data access.

Building Data Networks

As the demand for more comprehensive (and interoperable) data sources grows, building and participating in data networks has become an increasingly sought-after route for accessing patient data in Europe. Data networks function as a best-of-both worlds approach between registries and target site access, bringing together multiple healthcare providers, research institutions, and sometimes even public health agencies to create a more integrated and extensive data-sharing environment. Unlike registries, which are focused on collecting and managing data on specific diseases or patient populations, data networks offer a broader, more dynamic framework that allows for the aggregation and analysis of diverse datasets across various conditions, populations, and regions. Additionally, data networks facilitate interoperability between different systems and institutions that may, on their own, adhere to disparate data models and care practices.

COVIDRIVE represents a prime example, a highly effective private-public partnership network to track SARS-CoV-2 infections, characterized by a master protocol and sponsor-specific protocol supported by 20 sites. COVIDRIVE relies on data abstraction into an EDC (electronic data capture) rather than ETL (extract, transform and load). Other examples of such networks include EHDEN (European Health Data & Evidence Network) and the IMI2 (Innovative Medicines Initiative) projects like the BigData@Heart consortium. EHDEN and IMI2 aim to facilitate large-scale data access and analysis by connecting various data sources across countries and regions.

Despite their potential, these networks also face significant limitations. Data interoperability remains a major challenge, as differing data formats, standards, and languages across Europe can hinder seamless integration. Site data models can range from unstructured data lakes to the Observational Medical Outcomes Partnership (OMOP) Common Data Model. However, data architecture is often not truly known until access is granted for feasibility, as OMOP-designated sites may still have idiosyncratic data features that require further harmonization. Additionally, the governance of these networks is complex, involving multiple stakeholders with varying priorities and concerns about data privacy and security – particularly with private-public partnerships. And, as previously discussed, identifying and accessing new sites for network growth is a resource intensive task – scalability remains a major challenge. While these networks are a promising development in the field of data access, they are still evolving and often require substantial investments in infrastructure and collaboration to function effectively.

Barriers to Access

While data registries and individual site access provide valuable starting points for data collection, they often fall short in addressing the complex needs of large-scale studies. For comprehensive research, particularly for understanding patient journeys or burdens of disease across countries and care models, these traditional routes can be insufficient. This is where building a network of sites becomes critical, but it also introduces a new set of barriers.

Regulatory Complexities

One of the most significant barriers to accessing local patient data in Europe is the complex regulatory environment – relevant for both register and site data acquisition. European medical data is among the most well-protected in the world due to the General Data Protection Regulation (GDPR). GDPR's stringent requirements around data handling, access, and sharing are essential for protecting patient privacy, but they can also make it challenging to share or access data across borders or even within the same country. Each EU member state can implement additional regulations, leading to a patchwork of rules that can vary widely. This inconsistency complicates efforts to aggregate or analyze data at a European level, with compliance often being time-consuming and resource intensive. Moreover, the fear of financial penalties and negative publicity for violating GDPR has created a cautious environment. Hospitals and investigators may hesitate to share data—even when legally permitted—due to the perceived risk. This is further complicated by cultural and ethical differences across Europe, where attitudes towards data sharing and privacy vary significantly, influenced by historical, social, and political factors.

Varied Healthcare Systems and Practice

Europe's healthcare landscape is heterogeneous, with each country operating its own healthcare system, often with substantial differences in both care practice and how patient data is collected, stored, and managed. Such disparities make it difficult to both collect and analyze patient data, as there are subtle differences in care delivery and coding practices that are easy to overlook, leading to incomplete cohort capture, outcome measurement errors, and inaccurate comparisons . This can be particularly problematic when trying to conduct large-scale studies that require comprehensive, cross-border data. Researchers must therefore have local clinical insights to account for these variations.

Data Localization Requirements

Depending on site infrastructure and governance, institutions may impose data localization requirements, mandating that patient data be stored and processed within national borders. These requirements are driven by concerns over data sovereignty and security but can create significant barriers to accessing data for cross-border research or healthcare initiatives. Data localization can lead to the duplication of infrastructure and resources, increasing the cost and complexity of data management. Additionally, these requirements can hinder the ability to conduct large-scale studies that require data from multiple countries, as researchers may be unable to access or combine datasets from different jurisdictions.

Variability in Data Quality and Standards

The quality and standards of patient data vary widely across Europe, reflecting differences in healthcare practices, technology adoption, and data governance. In some countries, electronic health records (EHRs) are widely used and well-structured, while in others, paper records may still dominate, or digital systems may be poorly integrated. This variability can make it challenging to obtain reliable, high-quality data for research or clinical purposes. Inconsistent data standards also complicate efforts to harmonize datasets from different sources, as data may need to be extensively cleaned or transformed before it can be used effectively.

Overcoming the Barriers

Addressing the challenges associated with building a multi-site, international data network required innovative approaches tailored to each specific barrier. In this section, we outline the solutions implemented to overcome regulatory complexities, heterogeneous healthcare systems and practices, data localization requirements, and variability in data quality and standards. Our goal was to develop a robust, scalable network capable of supporting the comprehensive data needs of large-scale studies.

Navigating Regulatory and Contracting Complexities

To tackle the stringent and varied regulatory requirements across Europe, we adopted a proactive and flexible approach:

Engagement with Local and Regional Experts: We worked closely with local investigators, as well as regional and national regulatory experts, to ensure our activities were fully compliant with local data access regulations. This collaboration was essential for navigating the complexities of GDPR and other country-specific laws, allowing us to align our processes with the expectations of both regulators and data custodians.

Protocolized Network for Standardized Approvals: Our network is fully protocolized, meaning that all studies to be conducted within the network are intended to follow standardized pathways for obtaining data access approvals. This structured approach will provide a clear and consistent framework, facilitating smoother approval processes that are familiar to ethics committees, ensuring data is accessed in a controlled and compliant manner.

Advance Planning and Time Buffers: Recognizing the potential for delays due to lengthy contracting and approval processes, we planned well in advance and built significant time buffers into our project timelines. This proactive planning allowed us to manage delays effectively without jeopardizing project deadlines.

Flexibility with Site-Level Requirements: We adopted a flexible approach when it came to site-specific data access requirements. Whether gaining direct access to a hospital's data warehouse, working in a virtual environment via cloud-based solutions, or physically working on-site, we were adaptable to the needs and constraints of each site. We also collaborated with other research groups that had already established relationships with sites, working with third party data warehouses when necessary.

Overcoming Heterogeneous Healthcare Systems

The heterogeneity of healthcare systems across Europe presented significant challenges in accessing and integrating data. To address this, we:

Expert Engagement

As with the regulatory strategy, frequent communication with local healthcare professionals (HCPs) and their data teams is crucial for understanding both clinical practices and unique coding structures. For example, in France, hospitals often use CCAM (Classification Commune des Actes Médicaux) & UCD (Unités Communes de Dispensation) codes for procedures and treatments differing significantly from the ATC (Anatomical Therapeutic Chemical Classification System) codes used in other regions (e.g., Italy, Spain). To further the complexity, some sites may use multiple code types across (and even within) specialties, which can impact accurate cohort capture and outcome measurement. To address this problem at scale, distributing standard questionnaire to clinical and IT staff enables quicker feedback for multi-site analysis. Such detailed understanding is vital for accurate data interpretation and integration across different sites. Regarding health care practice, engagement is again critical, throughout the entire research term. Even with a robust data feasibility process, variances within and between sites will continually emerge and often require local insight to rectify.

Working with Data Localization Requirements

Requirements Data localization requirements pose a significant barrier to cross-border research. To overcome this, we:

1. Localized Data Hubs: Established localized data hubs in countries with strict data localization laws. These hubs allowed data to be stored and processed locally, in compliance with national regulations, while still being part of the broader network.

2. Federated Data Architecture and Analysis: Perhaps of most importance, we implemented a federated data model that enabled analysis to be conducted locally, with only the aggregated results being shared across borders. This approach minimized the need to move raw data between countries, thus reducing regulatory hurdles and ensuring compliance with data localization laws. Moreover, this approach is in line with upcoming data regulation (citation?) which will further prohibit pooling site data and require additional, often cost and time prohibitive, contractual agreements. Site variation (even within national borders) meant that our team was able to access the raw data warehouse of one large hospital network in France but only pass code to the IT staff at another site, who in turn passed over aggregate results. This was a frequent occurrence and forced our team to create new methods of analysis and data presentation.

Ensuring High Data Quality and Standards

The variability in data quality and standards across Europe posed a significant challenge, requiring a focused effort to ensure consistency and reliability in our data collection and analysis. We addressed this through several key strategies:

Landscape Analysis: To prioritize sites effectively, we developed a site prioritization matrix based on a comprehensive landscape analysis. This matrix considered factors such as the quality of available data, the consistency of coding practices, and the alignment of site capabilities with the study's objectives. This approach allowed us to identify and prioritize sites that were most likely to provide high-quality, reliable data.

Standardized Data Collection Tools and Natural Language Processing: We deployed standardized data collection tools and protocols across all participating sites. This standardization was critical for reducing variability and improving data consistency. In addition, we leveraged natural language processing (NLP) technologies to harmonize data from diverse sources, enabling us to integrate different coding systems and clinical terminologies into a unified framework. This ensured that the data collected from various sites could be compared and analyzed reliably.

Clinical Consultation and Training: We provided extensive training and support to local teams on data management best practices. This included guidance on the use of standardized data collection tools, ensuring compliance with data standards, and emphasizing the importance of accurate data entry and reporting. By involving clinical experts in the consultation process, we tailored our training programs to address the specific needs and challenges of each site, thereby enhancing the overall quality and reliability of the data collected.

These solutions enabled us to overcome the significant barriers associated with building a multi-site, international data network, ensuring that we could deliver the comprehensive, high-quality data needed for large-scale research studies. By sharing these learnings, we hope to provide valuable insights for other organizations facing similar challenges in the evolving landscape of healthcare data access.

Conclusion:

As we move towards a more interconnected and harmonized European health data ecosystem, the strategic approaches to timely and relevant data access presented here will become ever more critical.

Moreover, with the full implementation of the EU HTA regulation by 2025, it's imperative that life sciences organizations refine their data strategies to adapt to and navigate the complexities of the European data landscape - Nordic registries no longer serve as complete data solutions. Successfully addressing these challenges will allow these organizations to not just accommodate new evidence standards, but to conduct more efficacious research that examines and analyzes patient populations in their own healthcare environments.

graticule™