The term “real-world data” is broadly applied and relates to diverse methodologies and approaches. Along the pathway to market access, data from outside of the clinical trial setting is considered a “must-have” to support compelling messages of product value, safety, and effectiveness. Regulatory bodies, like the National Institute for Health and Care Excellence (NICE) in the UK and the Institute for Quality and Efficiency in Health Care (IQWiG) in Germany have recognized that randomized controlled trials (RCTs), although the gold standard for assessments of efficacy, do not provide any information about drug effectiveness in the real-world care setting. Also noteworthy is the participation of NICE in a pan-European initiative to develop a uniform framework for the use of real-world evidence (RWE), the first results of which have been summarized in a recently published technical support document.1 Other initiatives include the IMI GetReal2, a consortium consisting of the European Medicines Agency (EMA), health technology assessment (HTA) bodies, academia, patient organizations, and pharma aiming to establish a framework to assess the relative effectiveness of medicinal products.

An International Society of Pharmacoeconomics and Outcomes Research (ISPOR) Real-World Task Force has defined real-world data quite simply as those data used for decision-making that are not collected in conventional RCTs and has identified six discrete sources.3 (see Table 1.)

Additional sources of real-world data such as social media and cloud-based wearable health technologies add further to an already complex plethora of data available for life sciences research.

Evidence generation planning for successful market access must also include a multidimensional, real-world data strategy.

The early delineation of key value messages and associated evidence requirements, alongside a structured review of existing data gaps, are critical first steps for successful market access. This stepwise evidence generation planning process will serve to both identify and prioritize the research activities of importance – including the determination of real-world data needs. Questions related to “which data” from “which sources” and “what methodology” to address priority research objectives are at the core of a tailored fit for purpose data strategy.4 The identification of optimal data sources and the robust derivation of meaningful outcomes, amidst the chaos of massive amounts of often fragmented snapshots of patient experiences accruing daily, can be a very complex exercise.

“The right data for the right research question” ... the availability and suitability of each potential real-world data source must be thoroughly evaluated.

Real-world data can be obtained from existing sources or registries, including commercial data sources such as health insurance records, other administrative sources, or electronic medical records. The numerous existing sources of data can be either regional or restricted to specific healthcare facilities (e.g., specific hospitals), nationally representative, or even multinational. They differ not only in their content, but also the quality of
Tailored de novo data collection studies can be designed to resolve data gaps, but structured feasibility assessments are paramount prior to study initiation.

In addition to a systematic appraisal of potentially suitable data sources, a thorough delineation of real-world data gaps and potential biases should be undertaken. In the context of multinational evidence generation activities, inevitably a mix of database analyses as well as de novo data collection studies across countries or regions will be required to achieve a robust

The data, sample sizes covered, inclusion and exclusion criteria applied, and the settings of care that are being covered (hospital setting vs. outpatient setting). In addition, each of the data sources is ruled by its own terms and conditions that define data access that need to be taken into consideration when designing an RWE strategy, as these have great implications on the implementation timelines. Figure 1 summarizes the main criteria used to assess the availability and suitability of potential data sources.

Figure 1. Overview of a systematic approach to database evaluation

Table 1. ISPOR Task Force Sources of Real-World Data

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplements to traditional registration RCTs</td>
<td>Collection of data alongside clinically focused randomized controlled trials (RCTs); data include patient-reported outcomes (PROs), healthcare resource utilization, direct medical, and direct costs</td>
</tr>
<tr>
<td>Pragmatic clinical trials</td>
<td>Large trials that aim to measure effectiveness in routine clinical practice; the design reflects variations between patients that occur in real clinical practice</td>
</tr>
<tr>
<td>Registries</td>
<td>Prospective observational cohort studies of patients with a specific disease and/or receiving a specific treatment involving prospective data collection</td>
</tr>
<tr>
<td>Administrative data</td>
<td>Retrospective data collected primarily for reimbursement but also contain diagnosis and procedure use and detailed information on charges</td>
</tr>
<tr>
<td>Health surveys</td>
<td>Designed to collect reports on health status and self-perceived well-being, healthcare utilization, treatment patterns, and healthcare expenditures from patients, caregivers, healthcare providers, or individuals in the general population</td>
</tr>
<tr>
<td>Electronic health records and medical chart reviews</td>
<td>Electronic data capture facilitates medical chart reviews (either prospective or retrospective) in the creation of datasets with longitudinal disease specific data at the patient level through data abstraction</td>
</tr>
</tbody>
</table>

A systematic assessment of data sources is paramount and criteria for evaluation are varied
evidence base that has been adapted to the needs of each market.

Prior to the final design and initiation of any de novo data collection initiative, a comprehensive feasibility assessment at potentially eligible study sites is essential. A vital component of a thorough real-world data strategy, a carefully planned site feasibility assessment will serve to mitigate study risks as well as to inform project planning, including robust estimates of patient enrollment rates. Areas of key focus should include:

- Practice size and the existence of any mandated treatment pathways related to the therapeutic area of study
- Number of eligible patients by subgroup of interest treated per week or per month
- Medical chart management infrastructure and availability of key variables, if design incorporates medical chart review methodology
- Site institutional review board (IRB) and contracting processes, including unique requirements and timelines
- Availability of study staff for research conduct

Figure 2. A comprehensive real-world data strategy encompasses diverse methodologies and prioritizes technology and innovation
Data standardization and advanced or technology-enabled data analytics will ensure faster time to robust outputs and data interpretation. A well-designed real-world data strategy can result in a multinational patient-level repository of real-world information that has accrued from a diversity of sources, including de novo data collection efforts. The utility of these custom repositories is greatly enhanced, however, when a common data model\textsuperscript{5,6} that serves to standardize data vocabularies and formats is implemented. Standardization is highly recommended because it allows for the pooling and rapid analyses of highly variable and disparate data which inevitably result from programs of real-world research, as well as the following additional benefits:\textsuperscript{7}: 

- improved efficiency, through reduced programming time,
- increased transparency as a result of “analytics democratization” and the opportunity to share coding algorithms,
- reproducibility of results across datasets, and
- “faster time to data” by leveraging automated data analytics tools, such as Evaytica.\textsuperscript{8}

A comprehensive data strategy can provide a framework (see Figure 2) not only for the organization and prioritization of data sources and study types optimally suited to address the research questions of interest, but also to encourage various stakeholders within and across life science companies to plan for greater and more effective use of real-world data.

For more information, please contact Dimitra.Lambrelli@evidera.com or Krista.Payne@evidera.com.

REFERENCES


\textsuperscript{5} Observational Medical Outcomes Partnership Common Data Model. Available at: http://omop.org/CDM. Accessed April 9, 2016.

\textsuperscript{6} Observational Health Data Sciences and Informatics (OHDSI) Data Standardization. Available at:http://www.ohdsi.org/data-standardization. Accessed April 9, 2016
