Turning Real World Data Into Real World Evidence

The Pitfalls to Avoid to Best Aid Decision Making

KMK

KMK Consulting, Inc.

Your Integrated Commercial Operations Partner

The Rising Importance of Real World Evidence

Real-world evidence (RWE) now plays a significant role in how the pharmaceutical industry brings a brand to market. It can be used in assisting clinical trial designs, reviewing pricing structure and marketing strategy, understanding competitive threats and even determining new formulations, to name a few. While traditional Randomized Controlled Trials (RCTs) remain the golden standard to demonstrate clinical efficacy, given the typical confined and sophisticated nature of trial design, RWE can be a complementary means of illustrating drug efficacy and cost-efficiency in different real-world settings. More broadly, these real-world evidences are generated to provide safety surveillance, adding value to the marketed product, as well as serving embedded reimbursement and regulatory purposes.

In a nutshell, RWE is any insights generated by **real-world data (RWD)** outside the traditional clinical trial's Venn diagram. RWD uses massive patient level datasets to determine interventions' effectiveness in real-world circumstances. The data should be strictly selected, validated and standardized before it can be analyzed as evidence in order to:

- Understand the patient journey
- Provide epidemiology information,
- Estimate the burden of illness, etc.

all of which consequently helps the pharmaceutical industry provide better healthcare service to address the needs of patients.

Understanding The Sources And Reliability Of Real World Data Sets

Different RWD is gathered from a wide variety of sources and generated from multiple types of data. However, at its very core, all RWD comes from patient activities in a real world setting; it is what we call patient-centric data. According to the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force[1], sources of RWD are classified as:

1. **Supplements to RCT:** Providing additional information alongside clinical focused RCTs, such as patient reported outcomes (PROs) and cost information, etc.
2. **Pragmatic Clinical Trial:** Prospective and randomized assignment with larger and more diverse populations, which are used to measure the costs in real-world settings.
3. **Registry:** Prospective cohort studies of patients with specific characteristics in common. Patients are always followed for a longer time. Registry can be designed for different purposes, such as understanding disease progression, monitoring the post-market safety profile, and assessing the cost-effectiveness of particular treatment etc.
4. **Administrative Data:** Also known as claim data, are collected for reimbursement purposes, which are submitted by healthcare providers to payers when a patient uses health services. It provides the diagnosis, procedure and cost information which can be used to estimate clinical and economics outcomes.
5. **Health Survey:** Designed to collect epidemiology data such as health status, health care utilization and health care costs, etc. from representative individuals in the general population. Survey data can be used to understand the generalizability of impact from certain treatments on the target population.
6. **Electronic Health Record:** EHR data are used to capture real time clinical and laboratory data. Health care providers routinely enter clinical and laboratory records for patients' care.

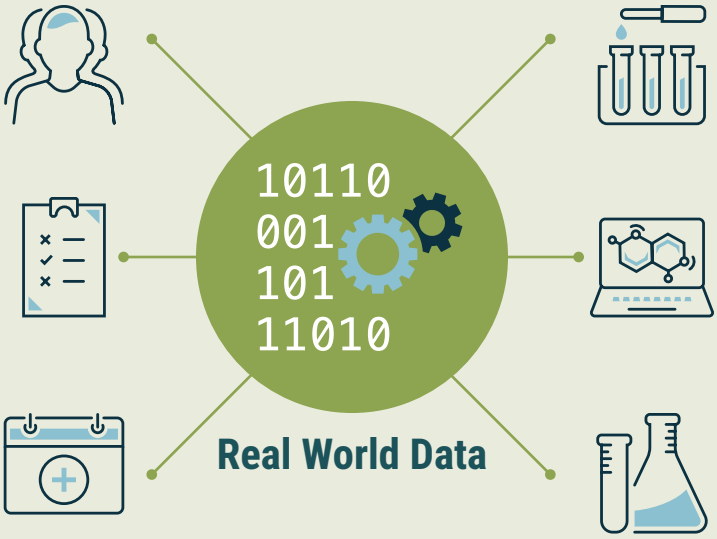
While traditional Randomized Controlled Trials (RCTs) remain the golden standard to demonstrate clinical efficacy, given the typical confined and sophisticated nature of trial design, RWE can be a complementary means of illustrating drug efficacy and cost-efficiency in different real-world settings.

Sources of Real World Data

Administrative Data
- Collected for reimbursement purpose
- Provide diagnosis, procedure and cost information

Health Survey Data
- Designed to collect epidemiology data
- Provide generalizability impact of treatment

Registry Data
- Prospective study of patients with specific characteristic in common



Pragmatic Clinical Trials
- Prospective and randomized trials with larger population
- Measure effectiveness in real world

Electronic Health Records and Medical Chart Review
- Collected for patient care purpose
- Provide real time clinical and laboratory information

Supplement to traditional RCTs
- Provided additional information alongside RCTs

Different sources of RWD depict the patient journey from different perspectives. A meaningful real world analysis requires thoughtful consideration of study designs and the nature of RWD sources. At the current moment there is no "jack-of-all-trades" data that fits all business needs. Each type of data has its own strengths as well as limitations. For example, pragmatic clinical trials, registry, and health surveys are prospective data sources which offer particular information and are suitable for certain types of studies or clinical trial questions. But the data collection process of a prospective study is economically costly and time consuming. Moreover, since the questions are already designed for a certain group of patients or outcomes, confounding or selection bias might exist if the collection process is not carefully used.

Given the large number of available RWD and the complexity of data structures, it has become one of the top priorities for researchers to select the optimal data sources to meet different business needs. Efforts and expertise should be brought to bear on the careful evaluation of the information each data set provides as well as the presence of potential bias and limitations.

Turning RWD into meaningful RWE is as edifying as it is challenging. One should develop or work with partners having hands-on experience with a wide array of domestic and international RWD sources. Such experience will lead to the best identification of the key data elements based on the study's research questions, thereafter selecting the most suitable RWD.

Possibilities For Turning RWD into Meaningful RWE

The following hypothetical scenario illustrates the expertise needed to find the best approach while conducting RWD analysis.

ABC Pharmaceutical Company launched a low-density lipoprotein (LDL) lowering drug XYZ two years ago. On the original label, the drug was intended for adults with high LDL cholesterol level. The sales team at ABC has received positive feedback from many patients who indicated they have experienced significant weight loss, and their body mass indices (BMI) are approaching healthier range. ABC wants to carry out a clinical trial to confirm such findings, and file to the Food and Drug Administration (FDA) for a new indication for XYZ of countering obesity due to high LDL cholesterol. To facilitate the design of the trial, they have contracted an external consulting firm to conduct a study.

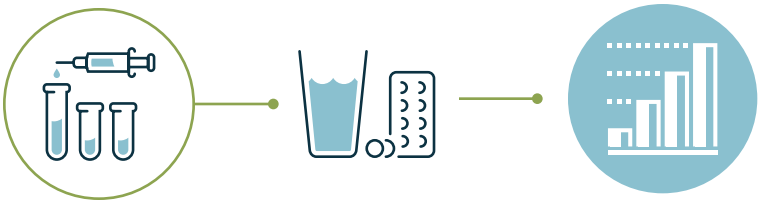
Here are some of the challenges of this hypothetical study, and how one might best proceed:

SITUATION/CHALLENGE:

To estimate a precise distribution of potential obese patients with high LDL cholesterol level.

SOLUTION AND IMPACT:

Before jump starting the project, we have to understand the goal of the study, which is to determine the obese patient group most susceptible to a weight loss, and how a trial can be best designed to recruit these potential patients. To access the potential trial applicants, it is important to understand the demographic impact on many available data sources. Most RWD data will reveal only a slice of the general obese population. For example, while a commercial insurance claims database may have coverage of more than one tenth of the general American population, these databases are usually very poor representations of the national population distribution. They typically skew toward a few states, age groups, and social groups. Similarly, commercial data from Electronic Health Records (EHR) systems are also inherently biased geographically and demographically. There is publicly available data that comes from surveys conducted by the Center for Disease Control and Prevention (CDC). The National Health and Nutrition Examination Survey (NHANES) data is likely the most relevant in depicting the national distribution of the obese patient. However, there is a quite a significant time lag for these survey data results as compared to the commercial insurance data (2-4 years vs. a few months), which makes it inapplicable for setting up this trial in a timely fashion.



SITUATION/CHALLENGE:

To accurately access the patient information associated with the desired outcome.

SOLUTION AND IMPACT:

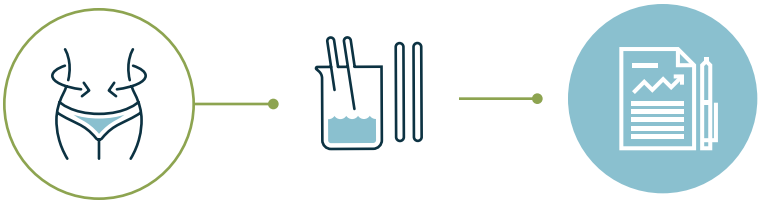
To validate if there is a positive correlation between events of LDL cholesterol level dropping with a dip in BMI, we need to first establish they are concurrent events, and then rule out any other possible reason for the BMI reduction. In this manner, we can be confident in reporting the BMI reduction has a positive correlation to the usage of the drug. None of this can be achieved without examining the actual numerical data of BMI and LDL cholesterol levels. International Statistical Classification of Diseases and Related Health Problems (ICD) codes and Current Procedural Terminology (CPT) codes are almost universally available natively in almost all United States based RWD databases. While ICD and CPT codes for obesity and cholesterol tests can be easily tracked down in commercial insurance claims data, they do not provide us with the appropriate information as they simply reflect billing of the test, not the result.

To extract the necessary data to facilitate the targeted analysis it is better to consult RWD data sources with an EHR origin, or a data source that has emerged from cross-pollination between insurance claims and EHR records. This is somewhat of a herculean task. As a fully competitive market, the United States healthcare providers have a variety of EHR systems to choose from. Most of these systems use different protocols to collect and store information. As a result, much of the commercial insurance RWD data was curated from a range of EHR systems. As the availability of such real world data is still in its infancy, data vendors do not necessarily always standardize results across different EHR systems, or even across the same EHR system with different software versions. It is not uncommon to find lab data left in its raw state. To illustrate this point, we have extracted LDL cholesterol lab test results from a leading commercial EHR database. In table 1 below, we have listed some of the possible units associated with the LDL cholesterol lab results which we have extracted from the database.

Some of these units are physically equivalent; others are of different magnitudes. There are also number ranges and units expressed that are most likely typos or hard to interpret correctly. It should be quite evident that this divergence could only be the result of data aggregation from a multitude of data sources. A novice mistake is to use the data as it is without a secondary data cleaning, or without even realizing there is such a need, which may lead to erroneous results.

Table 1: Units Associated With LDL Cholesterol Lab Results (from a leading commercial RWD source)

x10^6/ul	m	mg/dl (calc)	cm
uu/ml	milligram per decili	mg/dl	cc
u/l	mh/dl	mg/24hr	calc
u	mg-dl	mg//dl	70-129
other unit	mgd/l	Mg	0-99
NULL	mg/ml	meq/l	0-100
no units	mg/gl	log u/ml	<130 mg/dl
nmol/l	mg/dl?	in g/dl	<130
ng/dl	mg/dl.	Hr	<=99
mmol/l	mg/dl (calc(	Fl	%
x10^6/ul	mm	mg/dl (calc)	cm
uu/ml	milligram per decili	mg/dl	cc



SITUATION/CHALLENGE:

To determine the effectiveness of treatment.

SOLUTION AND IMPACT:

Even with all the data organized and ready, to identify and quantify a positive correlation between pre-/post-treatment cholesterol levels to BMI indices requires more than a just simple statistical analysis. How do we make sure the dropping LDL cholesterol level is the actual cause leading to the weight loss? Are the numbers sufficient to make any statistical inference? Some researchers may find it delightful to report positive findings, and quickly jump to the conclusion. After all, numbers make the best argument. But, in this case, do we have enough confidence that is indeed the case? More often than not, a seemingly scientific interpretation of statistics leads to incorrect conclusions[2]. A common root cause for statistically insignificant conclusions is a limited sized analytical patient cohort. To balance a thoroughly designed cohort selection criteria and create enough sample size to make valid statistical claims can be a tedious task, but one that cannot be overlooked in producing meaningful real-world evidence.

Conclusion

Along with the staple RCTS, RWE has come a long way in proving its value, becoming widely accepted as a measurement in the decision-making process. Despite the enormous information RWD generates, limitations and concerns about how to wisely use these evidences also exist. Challenges like potential selection bias and data quality issues inherited with RWD are well recognized by the industry. RWD can be a high-priced burden without the proper skills and experiences to translate those raw numbers into meaningful insights.

References:

1. Garrison LP Jr., Neumann, PJ, Erickson, P, Marshall, D, Mullins CD. Using Real-world Data For Coverage and Payment Decisions: the ISPOR Real-world Data Task Force Report, Sept.-Oct. 2007, https://www.ncbi.nlm.nih.gov/pubmed/17888097, (Accessed February 2019).
2. Huff, D., & Geis, I., How to lie with statistics. New York: Norton, 1954,

Authors

About Huanxue Zhou

Huanxue Zhou is Director of Health Economics & Outcomes Research (HEOR) at KMK Consulting, Inc. She obtained her Master's Degree in Statistics from Lehigh University. In her current role, she trains and leads HEOR analysts to execute HEOR studies to support clients. She is responsible for understanding clients' needs and ensuring the successful implementation of assigned projects. She provides consultation on appropriate research design, statistical analysis and interpretation of results. In the past seven years, she has worked with clients to develop many conference posters and peer-reviewed publications resulting from the execution of high-quality real-world data analyses in various therapeutic areas using a variety of observational databases.

About Bob Tian

Bob is a senior analyst at KMK Consulting, Inc. He joined the HEOR group at KMK in 2015. Since then he has been working with multiple large US and foreign real-world data projects, developing analytical methods for various goals. Bob has a Masters Degree in Industrial Engineering from ISyE at Georgia Institute of Technology.

KMK Consulting, Inc.

KMK Consulting, Inc. is a full-service consulting firm specializing in commercial operations support to the life science industry. Since our inception in 2000, KMK has grown to have more than 120 full-time employees, providing analytical support to clients on-site, as a project, or as SaaS that helps drive business decisions and improve the efficiency and effectiveness of commercial analytics and sales operations. We eliminate complexities for commercialization leaders by integrating:

- Accurate Marketing and Sales Analytics
- Market Research
- Sales Ops Software
- RWE/Health Economic & Outcome Research

The KMK HEOR team has extensive experience working with different RWD sources to solve assorted business questions. Our expertise in real-world evidence studies enables us to choose optimal data sources and to conduct sophisticated analysis to overcome data limitations. We are meticulous in how we balance selection bias and data quality issues in the design of a study, as well as how we conduct the analytical process. We always subject both the input data and the outcome to the highest level of scrutiny. Our dedicated team will turn the data into insights decision makers need ultimately leading to better efficiency.

KMK Consulting, Inc.

23 Headquarters Plaza, North Tower, 7th Floor | Morristown, NJ 07960

info@kmkconsultinginc.com | **kmkconsultinginc.com** | Phone: 973-536-0700 | FAX: 973-536-0702