Optum

Using real-world data to close the rare disease data gap

Part 1: Using real-world data to close the rare disease data gap

How do you define a rare disease? That depends on where you live. In the United States, a disease is considered rare if it affects less than 200,000 individuals. The European Union classifies a disease as rare if it affects less than one person in 2,000. In Japan, the cutoff is one person in 2,500. Despite these differences in definition, there is one very important aspect of rare diseases that everyone can agree on: researching and developing treatments for them is uniquely challenging.

Many hurdles

It's true that progress has been made since the Orphan Drug Act of 1983 was enacted to provide companies with incentives to develop rare disease medications, but there still remains a huge deficiency. It's estimated that there are 7,000 distinct types of rare and genetic diseases, which affect about 1 in 10 people in the United States.[1] However, 93%–95% of rare diseases lack an FDA-approved product for treatment.[2] Meanwhile, a study of clinical trials in the U.S., the EU and Japan found that only 16% of rare diseases had any related clinical trial activity.[3] (See the appendix at the end of this paper for a look at where pharma is investing its resources in the rare disease market.)

The numbers above are not surprising. Even with the Orphan Drug Act, there are many daunting barriers that stand between researchers and results, including:

- The natural history of rare diseases is often poorly understood.[4]
- Knowledge of the pathophysiology and clinical manifestations of these diseases over time is frequently incomplete.[4]
- For many rare diseases, well-characterized efficacy endpoints appropriate for the disease are not available.[4]
- Under-representation of genetic diseases and other types of rare diseases in health care coding systems due to their individual rarity. This contributes to a lack of ascertainment and recognition of their importance for health care planning and resource allocation and prevents clinical research from being performed.[5]

What are real-world data?

Real-world data (RWD) are data gathered in real-world settings, as opposed to data generated in controlled settings in clinical trials. Claims and EHR data are the foundation of RWD because they provide insight into a patient's record of health care utilization and the associated costs through encounters such as:

- Clinical assessments
- Diagnostic tests
- Labs
- Diagnoses
- Interventions

Using natural language processing to dig deep into patient records and provider notes is especially valuable in providing a better understanding of the progression of a disease and patients' responses to therapies.

Real answers

All of this creates a huge need among researchers to bolster the information available about rare diseases and the people they impact. They can do this by adding de-identified, real-world data (RWD) featuring claims and electronic health record (EHR) information into the equation. Such RWD can either be used to supplement an existing registry or help researchers find patients for which no registry currently exists. Certainly, this would be a welcome benefit to any researcher, but it's just a glimpse of the many advantages of using RWD.

By capturing data all along the continuum of care, RWD facilitates longitudinal outcomes research, development of patient maps and clinical patient profiling for commercial teams. RWD include data from clinical appointments, labs, diagnostics and assessments, data from diagnosis and treatment encounters, and post-surgical care data. This end-to-end view allows researchers to gain invaluable insights about:

- Risk factors for — and precursors to — developing a disease
- Treatment selection and line of therapy
- Short- and long-term clinical outcomes, including progression, response, remission, relapse, recurrence and death
- Complications and adverse events that require supportive care

When put all together, this can produce a remarkably clear picture of a treatment's effectiveness and a patient's experience.

Multiple uses

Since RWD can provide a robust, longitudinal view of a patient's path from the start, it can be used by stakeholders throughout the health care system. For example, the FDA is using RWD and the real-world evidence (RWE) that comes from that data to keep an eye on post-market safety and adverse events and to make regulatory decisions. The health care community is using RWD to make coverage decisions and to develop guidelines and decision support tools for use in clinical practice. And medical product developers are using RWD and RWE to support clinical trial designs and observational studies to generate innovative, new treatment approaches.[6] Increasingly, RWD are also being used for label expansions and to demonstrate to payers and providers the value of products for clinically precise groups of patients.

This all creates a remarkable opportunity for researchers and product developers in the rare disease space, especially since the collection and curation of data is far less costly and time-consuming than the execution of organic, randomized control trials. By seeking out RWD and embracing it in their work, researchers can fill the rare disease data divide and smooth the path to developing effective and profitable drugs for the rare disease marketplace.

What to look for in real-world data

If you're interested in putting real-world data to work, make sure you use high-quality data. That means using data that:

- Is sourced from a large and diverse group of providers
- Includes relevant patient populations across demographics, geography and disease states
- Is properly de-identified and directly sourced from providers
- Includes documented provenance and automated mapping of variables
- Includes data from clinical encounters in a wide range of specialties like dermatology, urology, pulmonology, gynecology, cardiology and neurology
- Includes data from broad therapeutic areas like type 2 diabetes, heart failure and COPD
- Is comprised of various types of data — such as closed claims, EHR and patient reported data — that can be easily integrated and used together
- Includes patients who are deterministically matched, such as patients whose EHRs are matched directly to their adjudicated medical and pharmacy claims

Where is pharma investing in the rare disease space?

To answer this question, we took two approaches:

1. We conducted a search of the rare disease database of the U.S. Food and Drug Administration, which generated the results in **Table 1**. The search was limited to those drugs that received approval between 2016 and 2019. The drugs approved during this period represent more than 30% of the rare disease drugs approved since the Orphan Drug Act was passed in 1983.[7] Oncology was by far the number one area of investment, with nearly 50% of the approved products.

2. Within the category of orphan diseases, there is a natural delineation between oncology and non-oncology orphan drugs. To drill down within those areas, we searched the literature to identify the top rare diseases for which pharma has developed therapies.

Table 1. FDA-approved orphan products by TA (2016–2019)

TA	# of approved products
Oncology	135
Hematology	31
Infectious disease	23
Neurological	17
Other	17
Respiratory	14
Autoimmune	10
Endocrine	9
Muskuloskeletal	9
Opthalmology	8
Dermatology	4
Renal	3
Cardiovascular	3
Liver	2
GI	1
	286*

*Represents more than 30% of drugs approved for orphan indications since the Orphan Drug Act passed in 1983.

Source: accessdata.fda.gov/scripts/opdlisting/oopd/index.cfm

Our search led us to a study which analyzed the leading clinical trial databases to identify the top rare diseases by clinical trial activity,[3] and a Pharma Intelligence report on activity in rare disease drug development.[8] We referenced Biomedtracker to benchmark the lists from these publications to identify the oncology and non-oncology rare diseases with the most drug development. The results are shown in **Table 2** and **Table 3**.

Table 2 – Top rare diseases by number of approved products (oncology)[9]

Type	**Estimated disease prevalence (/100,000)	Approved orphan products	Pipeline
Multiple myeloma	11.9 P^E	13	31
Acute lymphoblastic leukemia	11.0 P^E	11	13
Chronic lymphocytic leukemia – NHL	48.0 P^E	11	8
Indolent non-Hodgkin's lymphoma (includes follicular lymphoma)	11.6 I^E	11	10
Acute myeloid leukemia	10.0 P^E	9	58
Cutaneous T-cell lymphoma – NHL	24.0 P^E	8	8
Ovarian cancer	49.0 P^E	7	31
Chronic myeloid leukemia	6.0 P^E	7	4
Hepatocellular carcinoma	15.0 P^E	6	21
Diffuse large B-cell lymphoma – NHL	43.0 P^E	5	17

Source: www.biomedtracker.com.

**Without specification, published figures are worldwide. An E indicates European data, P indicates prevalence data and I indicates incident data.

Table 3 – Top rare diseases by number of approved products (non-oncology)[9]

Name	**Estimated disease prevalence (/100,000)	Approved orphan products	Pipeline
Cystic fibrosis	7.4 P^E	9	17
Hemophilia A/B	7.7 P^E	8	18
Pulmonary hypertension	3.0 P^E	7	16
Growth hormone deficiency	1.0 P	5	6
Amyotrophic lateral sclerosis	3.85 P	3	25

**Without specification, published figures are worldwide. An E indicates European data and P indicates prevalence data.

Part 2: The promise of EHR data in the study of rare disease

The need for more information

In part 1 we outlined the unmet potential of real-world data in rare diseases.[10] The need for more information about rare diseases can be seen in the slow progress made to understand their origin, progression and the effectiveness of new therapeutic options. Even with progress in new therapeutics since the Orphan Drug Act was passed in 1983, more than 90% of the known rare diseases are still without any known cures. Rare diseases come with high mortality, hence the urgent need to speed drug discovery and clinical development. An analysis of 350 rare diseases indicated that 27% of patients would not survive past their first year of life.[11] We searched the literature to identify the most commonly studied rare diseases in both the oncology and non-oncology conditions.[12,13] We then analyzed drug development activity for these conditions along with the prevalence of these diseases.

Table 1 – Top rare disease by number of approved products (oncology)

Type	*Estimated disease prevalence (/100,000)	**Approved orphan products	*Pipeline
Multiple Myeloma	11.9 P^E	13	31
Acute Lymphoblastic Leukemia	11.0 P^E	11	13
Chronic Lymphocytic Leukemia	48.0 P^E	11	8
Indolent Non-Hodgkin's Lymphoma (includes Follicular Lymphoma)	11.6 I^E	11	10
Acute Myeloid Leukemia	10.0 P^E	9	58
Cutaneous T-Cell Lymphoma – NHL	24.0 P^E	8	8
Ovarian Cancer	49.0 P^E	7	31
Chronic Myeloid Leukemia	6.0 P^E	7	4
Hepatocellular Carcinoma	15.0 P^E	6	21
Diffuse Large B-Cell Lymphoma – NHL	43.0 P^E	5	17

*Without specification, published figures are worldwide. An E indicates European data. P indicates prevalence data and I indicates incidence data.

**Source: biomedtracker.com

Table 2 – Top rare disease by number of approved products (non-oncology)

Name	*Estimated disease prevalence (/100,000)	**Approved orphan products	Pipeline
Cystic Fibrosis	7.4 P^E	9	17
Hemophilia A/B	7.7 P^E	8	18
Pulmonary Hypertension	3.0 P^E	7	16
Growth Hormone Deficiency	1.0 P	5	6
Multiple Sclerosis	309 P	5	2
Juvenile Rheumatoid Arthritis	4.2 P^E	5	0
Amyotropic Lateral Sclerosis	3.85 P	3	25
Idiopathic Pulmonary Fibrosis	11.5 P^E	2	22
Huntington's Disease	2.7 P^E	2	10
Malaria	3.0 P^E	2	4

*Without specification, published figures are worldwide. An E indicates European data and P indicates prevalence data.

**Source: biomedtracker.com

The dilemma for researchers

Rare disease research comes with well documented challenges. Small sample sizes are the norm.[14] Clinical trial recruitment has historically been difficult in rare adolescent and adult cancers.[15] Many trials fail to recruit enough patients to complete the study. A cross-sectional analysis of randomized clinical trials studying rare disease registered in a clinicaltrials.gov study identified the frequency of completed and noncompletion of studies of rare diseases. Of the 659 qualifying clinical trials, 199 (30.2%) were discontinued, and low patient enrollment (n = 64, 32.1%) was the most common reason for a study being prematurely terminated.[16] This challenge of small sample sizes impacts the ability to characterize both the effectiveness and safety of these rare disease agents. These challenges delay getting these therapies to those patients who need them the most.

How Optum can help

Electronic health records (EHRs) hold the promise of yielding research-ready data that may support study and approval of drugs to treat these conditions. The Optum de-identified EHR dataset includes more than 100 million longitudinal EHR lives sourced from 57 distinct sources of EHR data. The data are representative of primary care, specialty clinics and the inpatient experience. Optum EHR data have breadth, depth and the longitudinality necessary for outcomes research, patient journey maps and clinical patient profiling for commercial teams. It has breadth across disease states and across the patient's interactions with the health system.[17]

Optum has identified the ICD-9 and ICD-10 codes for rare diseases and the newly diagnosed and active patients for each cohort. For rare cancers, the corresponding CPT® codes retrieved commonly found mutations and biomarkers affiliated with the condition.

Table 3 — Rare cancer counts in Optum EHR database by active patients

Rare cancers	ICD-10 codes	ICD-9 codes	Total diagnosed	Newly diagnosed	Active patients	Biomarkers and gene mutations	Counts
Ovarian Cancer	C56.1, C56.2, C56.9	183	61,229	46,754	39,228	CA125: 86304, 81500, 81503, BRCA1, BRCA2: 81162, 81163, 81164, 81165, 81166, 81167, 81211, 81212, 81213, 81214, 81215, 81216, 81217, 81432, 81433, KRAS: 81275, 81276, 81210	18,293
Hepatocellular Carcinoma	C22.0	155	36,053	29,699	17,330	TERT, TP53, and CTNNB1, AXIN1CTLA-4, PD-L1, VEGFR serum AFP: 82105	11,668
Multiple Myeloma	C90.0	203.00, 203.01, 203.02	34,145	21,503	11,667	MYC, 17P DELETION, TP53, ATM, ATR gene, IRF4, EGR1 gene, MAPK pathway; KRAS, NRAS and BRAF gene: 81275, 81276, IGH: 81261, 81262, 81263, 81264	1,671
Chronic Lymphocytic Leukemia	C91.1	204.10, 204.11, 204.12	36,095	18,595	10,061	BTK: 81233 PLCG2: 81320 IGHV, TP53, 17p, CD38, ZAP70, CD49d	431

Table 3 — Rare cancer counts in Optum EHR database by active patients —continued

Rare cancers	ICD-10 codes	ICD-9 codes	Total diagnosed	Newly diagnosed	Active patients	Biomarkers and gene mutations	Counts
Follicular Lymphoma	C82.0, C82.1, C82.2, C82.3, C82.4, C82.5, C82.6, C82.7, C82.9	202.0X	16,482	10,610	5,387	BCL2, BCL6, MUM1 (search by names) EZH2: 81236, 81237 MYD88: 81305 IGH: 81261, 81262, 81263, 81264, T cell antigen receptor: 81340, 81341, 81342	513
Acute Myeloid Leukemia	C92.0, C92.8, C92.6, C92.5, C94.2	205.00, 205.01, 205.02	13,512	9,850	5,017	CEBPA: 81218 NPM1: 81310 FTL3: 81245, 81246 KIT: 81272 BCR/ABL1: 81206, 81207, 81208 RUNX1, ASXL1, TP53	2,115
Chronic Myeloid Leukemia	C92.1, C92.2	205.10, 205.11, 205.12	11,126	6,500	3,470	BCR/ABL1: 81206, 81207, 81208, CD34, CD33, ARHGEF1, TET2, ASXL1	2,562
Acute Lymphoblastic Leukemia	C91.0	240.00, 204.01, 204.02	8,563	5,056	2,363	CR/ABL 1: 81206, 81207, 81208 TPMT: 81335	705
Diffuse large B-Cell Lymphoma	C83.3	200.7X	5386	3220	1324	BTK: 81233, CD20, CD10, MYC, BCL2, BCL6, MUM1 (search by names), EZH2: 81236, 81237 MYD88: 81305 IGH: 81261, 81262, 81263, 81264	732
Cutaneous T-Cell Lymphoma — NHL	C84.8, C84.5 C84.4	202.7X	1,857	1,385	676	SAMD1, ATM, TP53, CDKN2A, CTLA4, CD28, MAPK pathway; KRAS, NRAS and BRAF gene: 81275, 81276, 81210	19

1. Time period for analysis in Optum EHR: 2015–2019
2. Newly diagnosed patients with a washout of 5 years
3. Active patients with at least 2 encounters a month apart
4. Biomarker and gene mutation counts are from newly diagnosed population

Table 4 — Non-oncology rare disease counts in Optum EHR database by active patients

Rare diseases	ICD-10 codes	ICD-9 codes	Total diagnosed	Newly diagnosed	Active patients
Pulmonary Hypertension	Q87.8, 127.2X, 127.0, N15.8	416.0, 583.89, 416.8	294,724	217,567	80,972
Idiopathic Pulmonary Fibrosis	J84.10	515	253,840	214,253	70,437
Multiple Sclerosis	G35, G37.8	340, 341.8	186,210	113,098	65,564
Growth Hormone Deficiency	E23.0, E34.3	253.2, 253.3, 628.1, 259.4	63,731	50,567	24,386
Hemophilia A/B	D68.4, D68.2, D68.1, D66, D67	286.7 286.3 286.2 286.0 286.1	69,701	58,843	19,300
Juvenile Rheumatoid Arthritis	M08.1, M08.2, M08.3, M08.4, M08.8	714.30, 720.0	40,659	26,343	11,352
Amyotropic Lateral Sclerosis	G12.21	335.2	10,611	9,700	5,892
Cystic Fibrosis	E84.0, E84.1, E84.8, E84.9	277.00 277.02 277.09 277.01 277.03	15,864	10,931	4,259
Huntington's Disease	G10	333.4	4,980	3,482	1,857
Malaria	351.0,B51.8, B51.9, B52.0, B52.8, B52.9, B53.0	084.9, 084.0, 084.8, 084.1, 084.5, 084.2, 084.3, 084.4, 084.6	4,561	4,237	565

1. Time period for analysis in Optum EHR: 2015–2019
2. Newly diagnosed patients with a washout of 5 years
3. Active patients with at least 2 encounters a month apart

Hope for researchers and patients

Optum EHR data contain insight that can aid rare disease research. It can be leveraged to conduct natural history studies, map the patient journey, quantify the existing treatment patterns, serve as a control arm for new drugs in development, and demonstrate the effectiveness of drugs being used off-label to treat rare disease. As a control arm or supplement to clinical trial, EHR real-world data may help speed the approval and expanded use of agents desperately needed by the patients with these serious medical conditions.

About Optum Life Sciences

Optum helps pharmaceutical, biotechnology and medical device companies successfully address product development and commercialization challenges. We combine unparalleled data and analytics expertise with comprehensive technologies and health care ecosystem knowledge to power modern health care. This breadth and scope of assets allows us to help life sciences companies pivot from the current view of a single patient episode, to a new, more comprehensive and inclusive approach to understanding how their medicines perform in the context of the overall health system.

Sources:

1. FDA. Rare diseases: Natural history studies for drug development, guidance for industry. fda.gov/media/122425/download. March 2019.
2. EveryLife Foundation. About Rare Disease. everylifefoundation.org/stat-act/about-rare-disease/.
3. Sakate R, Fukagawa A, Takagaki Y, Okura, H, Matsuyama A. Trends of clinical trials for drug development in rare diseases. Current Clinical Pharmacology. ncbi.nlm.nih.gov/pmc/articles/PMC6327118/. August 2018;13(3):199-208.
4. FDA. Rare diseases: Common issues in drug development, guidance for industry. fda.gov/media/119757/download. January 2019.
5. Aymé S, Bellet B, Rath A. Rare diseases in ICD11: Making rare diseases visible in health information systems through appropriate coding. Orphanet Journal of Rare Diseases. 2015;10(35). doi:10.1186/s13023-015-0251-8.
6. FDA. Real world evidence. fda.gov/science-research/science-and-research-special-topics/real-world-evidence. Last updated March 23, 2020.
7. FDA. Search orphan drug designations and approvals. accessdata.fda.gov/scripts/opdlisting/oopd/index.cfm.
8. Shin D. A bird's-eye view of the rare disease landscape. Pharma Intelligence, Informa PLC. pharmaintelligence.informa.com/~/media/Informa-Shop-Window/Pharma/Files/Reports/Rare-Diseases-Report_LR.PDF. January 2017.
9. Prevalence of rare diseases: Bibliographic data, Orphanet Report Series, Rare Diseases collection, January 2020, Number 1.
10. Hussain U. Using real-world data to close the rare disease data gap. Optum. June 2020.
11. Vickers P. Challenges and opportunities in the treatment of rare diseases. Drug Discovery World. Spring 2013. ddw-online.com/therapeutics/p211490-challenges-and-opportunities-in-the-treatment-of-rare-diseases.html.
12. Sakate R, Fukagawa A, Takagaki Y. Trends of clinical trials for drug development in rare diseases. Current Clinical Pharmacology. August 2018; 13(3): 199-208. ncbi.nlm.nih.gov/pmc/articles/PMC6327118/.
13. Shin D. A bird's-eye view of the rare disease landscape. Pharma Intelligence, Informa. January 2017.
14. Mitani AA, Haneuse S. Small data challenges of studying rare diseases. JAMA Netw Open. 2020;3(3):e201965. doi:10.1001/jamanetworkopen.2020.1965.
15. Burke M, Albritton K, Marina N. Challenges in the recruitment of adolescents and young adults to cancer clinical trials. Cancer. Volume 110, Issue 11. December 2007. Pgs 2385-2393. acsjournals.onlinelibrary.wiley.com/doi/full/10.1002/cncr.23060.
16. Rees CA, Pica N, Monuteaux MC, Bourgeois FT. Noncompletion and nonpublication of trials studying rare diseases: a cross-sectional analysis. PLoS Medicine. 2019;16(11):e1002966.w
17. Argentinis E. Seizing the golden opportunity to create an enterprise data strategy. Optum. 2019.

Let us help you connect the dots

From data to insight to action, we catalyze innovation and commercial impact.

Visit: **optum.com/life-sciences**

Optum

optum.com