Tools and Technology for Advancing Rare Disease Research and Drug Development

The Role of Emerging Science, Technology and Real-World Data

Contributors
Frederick Derosier, DO, Executive Director, Rare Disease and Pediatrics Team

Rapidly evolving science has led to a deeper, more definitive understanding of the biology of numerous rare diseases, and helped identify tangible targets for the development of novel therapies. However, despite this progress, much uncertainty exists around the progression and clinical course of many rare diseases. To address this issue, Labcorp is exploring novel methods to construct and exploit longitudinal real-world data sets in combination with advanced approaches for data analysis.

This white paper discusses our initial efforts at virtual mapping of natural histories, the application of predictive modeling to better understand comorbidities and disease progression, and examines the use of genotype-phenotype linkage using longitudinal real-world data sets. By uniting disparate sources of data from Labcorp and third parties, we hope to improve how we diagnose patients, design and conduct trials and enable the development of more treatment options for people living with rare diseases.

Recognizing the Challenges for Patients, Providers and Drug Developers

Rare diseases are often underdiagnosed or misdiagnosed. Patients typically consult multiple physicians and undergo extensive testing over many years before an accurate diagnosis is reached. This uncertainty also affects the drug development process where sponsors face challenges identifying and recruiting patients with rare diseases for trial participation, locating physicians with disease and/or trial experience and de-risking clinical development.

How can we address these key challenges for both patients and drug developers? One approach to bridge this gap explores the use of longitudinal, objective laboratory data and physician diagnostic coding from the patient care setting.

Leveraging Data from Laboratory Testing and the Voice of the Patient

Nearly 150 million people in the U.S. have had at least one laboratory test processed by Labcorp, which has amassed more than 30 billion test results, including genotype data, from the more than 5,000 diagnostic assays it offers. Beyond routine testing, results may include tissue, genomic and/or proteomic data; these results are linked to an ordering physician, ICD code and geographic location along with patient age and gender, and, in some cases, weight and vital signs such as heart rate and blood pressure.

These test results represent more than raw data values. Labcorp has also considered how to support medical research by providing patients undergoing testing with the option to receive notifications about clinical trials potentially relevant or of interest to them, as well as to participate in surveys. To date, more than 330,000 people have provided consent to be contacted by Labcorp. For patients with medical problems where therapy options are limited, this process provides a benefit by improving awareness of clinical trials that may offer another path for treatment.

A portion of those people who have provided consent to be contacted have also provided feedback via survey about their willingness to join a study and reasons why they would not join a study: for example, by specifying their upper limit for travel time, number of visits, or the duration of a study visit. This "voice of the patient" data feeds into the Labcorp Patient Intelligence portal, which can be used to better understand patients' mindsets about joining a study and to help a sponsor design a more appropriate, patient-centric protocol.

Addressing Privacy Concerns While Identifying Rare Disease Patients

Labcorp has strict privacy protections in place. Labcorp strictly limits who can access laboratory data and further ensures that any use or disclosures of patient data are done in compliance with the Health Insurance Portability and Accountability Act (HIPAA) and other applicable privacy laws. Labcorp Drug Development, as a contract research organization (CRO), only works with de-identified patient data and must collaborate with Labcorp Diagnostics colleagues to perform outreach to contact physicians, or patients, for potential study referrals to a Labcorp Drug Development-supported trial.

For example, Labcorp has already put this referral practice into place with a unique solution called the Physician Referral Network. Using the Labcorp integrated oncology database, a Labcorp cytogeneticist examines anonymized data from patients who underwent an esoteric Labcorp test for a specific translocation or insertion.

Based on the findings of the test, the cytogeneticist determines the likelihood that the tested patient would have the rare cancer and qualify for the study. This information is passed on to Labcorp representatives who can reach out to the ordering physician and provide information about the clinical trial and the nearest study site. After consideration, the physician may or may not then refer their patient to the study. This collaboration allows Labcorp to expand its approaches to patient recruitment while protecting the privacy of the potential trial participant.

Examining Populations of Interest with ICD Codes

Beyond finding physicians who treat particular types of patients, test results can be aggregated into longitudinal real-world data sets that document the evolution of laboratory testing over time for both individual patients and populations of patients. Populations of interest are compiled by using a predefined combination of laboratory testing and/or ICD codes from Labcorp data. Physicians who order laboratory testing have to provide justification for the testing based on the diagnosis or identified symptoms. ICD codes, therefore, provide a window into physician perspectives regarding a patient's clinical issues and presentation at the time testing is ordered. By analyzing ICD codes, a researcher may get insights into populational comorbidities, their evolution and disease phenotypes.

ICD codes may be analyzed for frequency, testing order and specific diagnosis or symptoms. If genetic tests were also ordered for the patients, this layer of data could be added on to evaluate genotype-phenotype linkages. Similarly, a data set could be divided into subgroups to look at ICD code association with specific mutations or laboratory parameters.

Investigating Genotype-Phenotype Data and Co-existing Conditions

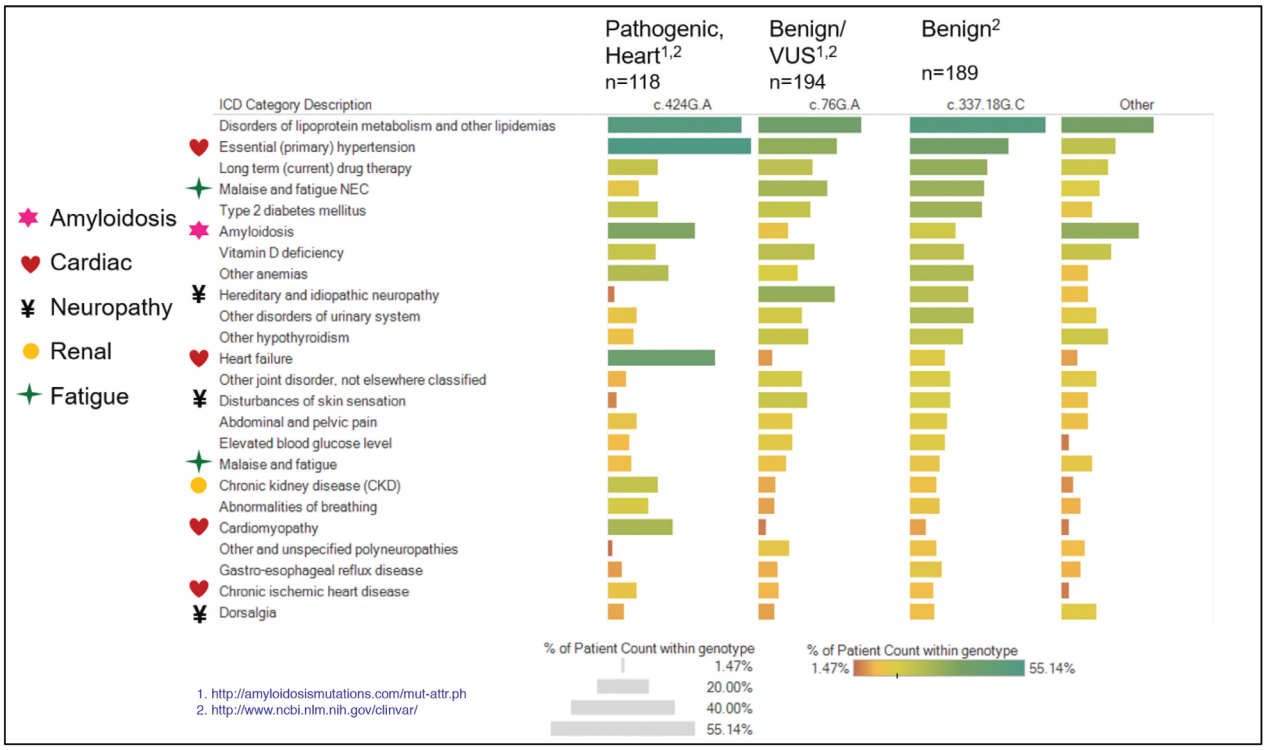
Analysis of longitudinal real-world data can help identify potential genotype-phenotype relationships and may provide supplementary information to the standard variant classification process determining mutational pathogenicity.

A brief case study outlines the application of real-world data for examining the relative pathology of transthyretin (TTR) mutations in amyloidosis (aTTR), where protein-misfolding results in destabilization of the normal TTR tetrameric structure into amyloid-forming monomers. In aTTR, these monomers deposit preferentially into cardiac tissue and/or nerves, and, to a lesser extent, into the kidneys and other organs, leading to tissue damage and progressive organ dysfunction. The predilection for deposition into one tissue or another is driven by the mutation, with some mutations being phenotypically more cardiomyopathic and others more neuropathic.

In the figure below, mutations are compared against co-existing conditions documented by the ICD codes compiled across the longitudinal data subset for individuals with the specific mutation of interest. One mutation is characterized as cardiomyopathic in expression; one is reported as a variant of uncertain significance (VUS) and the third as benign. However, the ICD-associated conditions from real-world data of individuals with these mutations suggests that the VUS may have a high percentage of neuropathic manifestations, while the mutation characterized as benign is associated with hyperlipidemia, hypertension and diabetes; approximately 20% of the individuals with this "benign" mutation in this data set have been assigned a diagnosis of amyloidosis by their treating physician.

While these data are not definitive, especially in the case of rare disease, seeing the associated conditions and clinical presentations helps paint a more detailed picture and may spur researchers to reconsider how a mutation is classified in terms of its pathogenicity. Patterns identified with longitudinal real-world data can also provide critical information about pathologic mutations and be particularly helpful to inform the development of advanced therapies, such as cell and gene therapies or gene editing.

Mutational-Associated ICD Coding Versus Pathogenicity



"Other" represents the rest of the mutations within this particular data set

Exploring the Patient Journey

As noted earlier, accurate diagnosis of rare diseases can be problematic, may involve multiple physicians, extensive testing—and may take many years. This journey to diagnosis, and hopefully effective treatment, can be quite challenging for patients and their families. The difficulty of the patient journey in rare disease is the exemplification of a number of issues including absent or incomplete natural history, heterogeneity of clinical presentation and lack of genotype-phenotype linkage in addition to physician unfamiliarity with a condition that may be exceedingly uncommon. Longitudinal real-world data sets may allow us not only to fill in some of these gaps, but can also be utilized to explore the patient journey and examine how a disease evolves over time from initial manifestations to diagnosis.

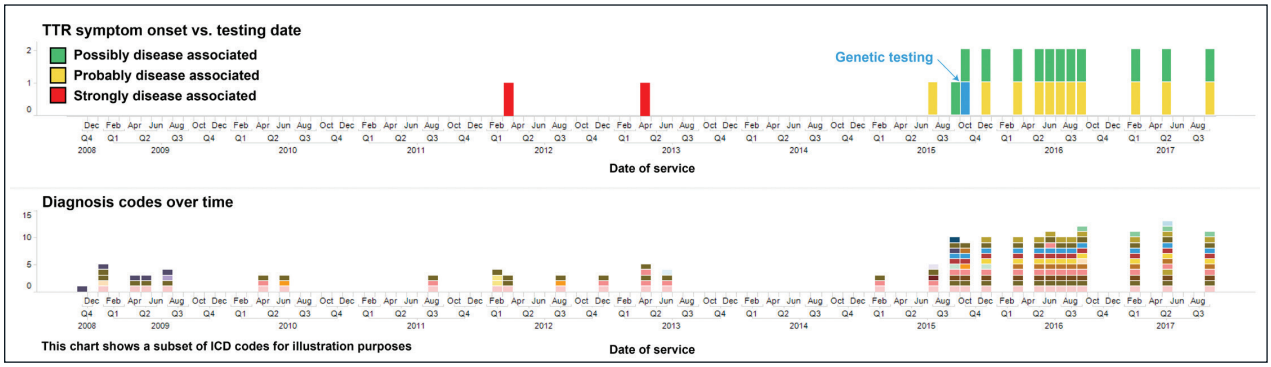
In the example below, all of the ICD codes associated with aTTR population data set were reviewed for known association and medical relevance to aTTR and color-coded: red is strongly associated; yellow, probably or moderately associated; and green, possibly associated with the diagnosis.

The image depicts how a male patient was definitely diagnosed on the basis of genetic testing as indicated by the blue box. It can be seen that the treating physician entered ICD codes for testing indicating conditions possibly and probably linked to aTTR just before and at the time that genetic testing was ordered for the definitive diagnosis, suggesting that these issues may have prompted the physician to consider aTTR in this patient.

Further, codes associated with aTTR appear with far greater frequency in this patient's record following the genetic testing, consistent with a definitive diagnosis. However, it is interesting to note that the patient had two ICD codes strongly associated with the presence of aTTR some 2.5 and 3 years before the genetic testing was performed, which can prompt several questions, such as:

- Did these earlier ICD codes represent a missed opportunity to diagnose the patient 3 years sooner?
- Are these conditions typical early manifestations of the disease and/or of this mutation specifically?
- Are there any biochemical alterations in the patient's laboratory testing that might be flags to pursue aTTR as part of the differential diagnosis?
- Is this a pattern consistently seen in other patients within the broader data set?

Patient Journey via ICD Codes



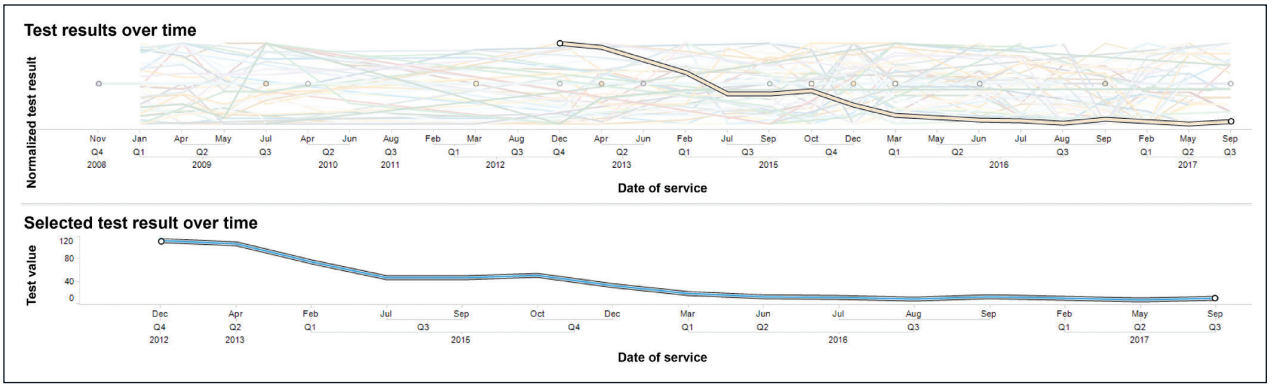
These are all areas for potential exploration with these types of longitudinal data sets. This particular individual suffered renal failure, a condition potentially resulting from TTR amyloid deposition in the kidneys.

We can use this example to ask larger questions, such as:

- How common is renal failure in this population and for this particular mutation?
- Is this presentation typical?
- What is the association with other known risk factors for renal failure, such as hypertension and diabetes?

We can clearly delineate a time course and degree of severity (rapid, and severe) for the renal failure in this individual, as shown in the figure below. The top bar shows eGFR highlighted within the collected longitudinal data of all laboratory testing for the individual, while the second bar segregates the event of interest for closer examination.

Patient Journey—Evolution of Comorbidity



Again, these are all questions that may be readily explored within this type of data, which can broaden our understanding of a disease and its progression pathways and identify symptoms and biochemical parameters which may help patients and physicians reach an earlier, more accurate diagnosis.

Using Machine Learning to Understand a Population

Elucidating the disease progression pathways can be enhanced through the application of predictive modeling with, for example, machine learning and artificial intelligence (AI) approaches. Machine learning also helps support data analysis at a population level to make inferences about a population and improve clinical development.

For example, if a pharmaceutical company was undertaking development of a drug with potential interactions in the kidney, their understanding, or lack thereof, of the likelihood of renal complications as a manifestation of the disease within their study could potentially impact patient safety, development timelines and even the success of the clinical development program.

At a broader level, machine learning could look at a disease as a whole with hypothesis testing. For example, with a longitudinal data set including genetic testing data for the population of interest, an AI tool, in theory, could identify mutations associated with decline in kidney function and determine the likelihood of developing the disease. These can be further categorized by age of onset, the time course and the severity of progression and even comorbid conditions. With this knowledge about the propensity of a group of individuals to develop renal failure, the sponsors can appropriately adjust their monitoring plan or their inclusion/exclusion criteria to mitigate the risk within the study.

Finding the Needles in the Haystack

Constructing large data sets may not always be possible for rare diseases, considering that overall size of a patient population can be quite limited. But even small data sets can be useful.

In a preliminary stage project, Labcorp researchers are examining an ultra-rare disease that has approximately 130 cases in published literature and an unknown prevalence; there are possibly 1,000 people in the world with the disease. However, as the disease may overlap in its clinical presentation with other conditions having a different etiology, and as none of the conditions have specific treatments which might warrant a more definitive diagnostic work-up, there is suspicion that the disease is under-recognized and under-diagnosed.

Consequently, Labcorp data was searched for a specific set of ICD codes and laboratory parameters that were felt to define a target population likely to harbor undiagnosed individuals with the disease. The data identified approximately 2,500 individuals who may have the disease and whose physicians will be asked to consider referring these patients for genetic testing. While the data itself is not diagnostically definitive, it enables a more efficient search to find "the needles in the haystack" and provides insight into how physicians evaluate different diseases with similar presentations.

The exercise of analyzing real-world data and investigating genotype-phenotype data helps support the key principles of precision medicine: to maximize efficacy and minimize side effects by matching the exact drug to the exact biological condition of the patient. Payers also benefit from these targeted efforts. Gene therapy, for example, is very costly. For insurance companies and payers, these types of data analyses would allow a more thorough evaluation of genotype-phenotype linkages and, in turn, allow a more informed decision about the appropriateness of a gene therapy treatment for a patient with a specific mutation, ultimately improving access for individuals who could truly benefit from the treatment.

Looking Ahead Toward a Collaborative, Knowledge-Sharing Environment

Whether informing a diagnosis, designing a study protocol or improving our understanding of a rare disease, real-world data has tremendous power and potential. With initiatives in place to leverage artificial intelligence and machine learning expertise using data from Labcorp, our team of rare disease researchers hope to draw actionable insights that influence drug development for rare diseases. Additionally, we are currently exploring methods for supplementing longitudinal data sets derived from Labcorp data with targeted data sourced from third parties. These expanded, longitudinal data sets have tremendous potential to improve how we diagnose patients, design and conduct trials and enable the development of more rare disease treatment options.

We plan to share our ongoing work in the public forum and add to the research communities' understanding of rare disease progression, identification and treatment. Ultimately, these collaborations will help improve patient-centric treatments at a global level and accelerate the development of personalized medicines for people living with a rare disease.

Learn more at drugdevelopment.labcorp.com

labcorp
Drug Development