The Power of Being Counted Report – A RARE-Xtra Podcast

Daniel Levine:

I'm Daniel Levine, and this is RARE-Xtra. For more than a decade, governments, nonprofits, and industry organizations involved in rare disease have stated as a matter of fact that there are 7,000 rare conditions, or estimated the number to be between 5,000 to 8,000.

The sources of these estimates are challenging to identify, given the circular nature of citations among groups repeating these figures. What's more, these estimates have remained static even though nearly 300 new rare genetic diseases are added to principle knowledge bases each year.

To develop a true count of rare conditions, RARE-X recently completed a research project that found the actual numbers approaching 11,000. We spoke to research lead and Alexion senior director of data science, Sebastian Lefebvre, vice president of patient experience for Alexion, Wendy Erler, and RARE-X CEO, Charlene Son Rigby, about the new paper, Be Counted, the significance of its findings, and what the implications are for rare disease patient communities. Seb, Wendy, Charlene, thanks so much for joining us.

Wendy Erler:

Thanks for having us.

Charlene Son Rigby:

Thank you.

Daniel Levine:

We're going to talk about a new report from RARE-X, that finds that the commonly used figure, that there are about 7,000 rare diseases, grossly understates the true number that are out there, why the actual number is approaching 11,000, and the significance of this.

Seb, let's start with you since you did the research behind this finding. And congratulations on the impressive work. I should also congratulate Kirk Lamoreaux, who was the lead author on the paper, and who like you, was diligent in doing this work. But let's start with the numbers. What did you find?

Seb Lefebvre:

So first, we start with the data source. We used both Orphanet, which is a European database, and OMIM, which is out of John Hopkins, as the two main sources of rare disease information. And we took the data as of December 11th, 2021.

By bringing this data together and removing duplicates and overlap, we got to 10,869 rare conditions. At a high level, we decided to focus on sensitivity. And as such, we ended up counting all disorders that have no subtypes. For example, if you take Batten disease, it has six subtypes like CLN1. In this case, we would count six subtypes for Batten disease.

Daniel Levine:

What's the argument for doing that? As I think of the therapies that are out there today for Batten disease, not that there are many, but usually they're targeting just one subtype.

Seb Lefebvre:

That's part of the rationale for counting the subtypes, essentially, which is whether it's genetic, whether it's a different age of onset, whether it's different symptoms, or emergent, or the pattern of inheritance—you got the variant from both of your parents, only one of them—that impacts the severity of some of those conditions. So, focusing on the subtypes is the way to go when you look at where we're heading with precision medicine.

Daniel Levine:

So, 10,869. Break that down for me. Help me understand what that number consists of.

Seb Lefebvre:

Yes. Let's first, just a high level, look at where the contribution for this number comes from. Orphanet brings you roughly, 6,000 disorders with an additional thousand subtypes. So, that takes you within the 6,000 to 8,000 ballpark. Now you add about 2000 rare genetic conditions for OMIM that do not overlap with the Orphanet set. And then you add one last chunk, which is the 2,500 rare genetic conditions that are essentially subtypes of an Orphanet entry, and that brings you over 10,000.

Once you look at this data set, 61 percent, roughly, have known a genetic underpinning. Twenty five percent are suspected genetic; so mostly OMIM entry, because OMIM is focused on genetic diseases. And then there's 13 percent that are just unknown from a genetic standpoint. And then, if you break it down further by classification, for example, 23 percent of them are developmental in nature. Nineteen percent don't have classification at all, and most of those are OMIM entries. Fourteen percent are neurology, and 5 percent are metabolic. Noteworthy, oncology only makes 3 percent of this set, and then 38 percent make up the rest of those diseases.

And I will leave you with one last way to look at this 10,000 plus diseases: 20 percent are definitely poorly defined, so they lack what we call phenotypic information. Basically, symptoms and biomarkers are just not available for those. Eighty percent of them do have some phenotypic information to some degree and that is important for the diagnostic part of this equation.

Daniel Levine:

You mentioned Orphanet and OMIM. For listeners who may not be familiar with these, what are they and why did you start there?

Seb Lefebvre:

Yes. Orphanet is your European database, and their job is really to bring together information about all those rare diseases that are seen in Europe, but also around the world. They've done a fantastic job at pulling together and classifying all those conditions, essentially.

Daniel Levine:

This is an authoritative database?

Seb Lefebvre:

Yes. It comes from cases that they see around the Orphanet clinical network.

Daniel Levine:

And OMIM?

Seb Lefebvre:

OMIM is out of Johns Hopkins and essentially that dataset or website is really focused on what they call Mendelian disease, which is genetic in nature. Anything that is inherited or has a genetic cause finds itself, to some extent, in OMIM.

Daniel Levine:

We keep hearing that there 7,000 rare diseases, that there are between 6,000 and 8,000 rare diseases. There has been some notable work by the Monarch Initiative that aligns quite closely with your findings. But how are these numbers derived and why is there this discrepancy?

Seb Lefebvre:

I think, mostly due to how we count rare disease, whether you group them up or start counting the subtypes. I think that's one key factor. But also, the way we name and describe the disease today. We describe them by the physician who first found the case, whether by the main symptoms or physical representation of the disease. We're not so much focused on the molecular cause, even though, as I mentioned earlier, the majority of rare diseases are genetic in nature. So that's another factor. And then lastly, on a monthly basis, new genetic diseases are being discovered. So, this number will continue to grow. What we have done, just like Monarch has done, is we just took basically a precision medicine viewpoint and focused on identifying the various patient population, so that we counted all the disease subtypes because their underlying molecular cause may vary; could be your inheritance, your gene, your age at onset, the way the disease progresses, the severity, and if, eventually, the treatment varies at the subtype level. And that's why we went down to the subtype, where I'm guessing, a lot of those publications around the 7,000, or 6,000 to 8,000 are more at the grouping level than the subtype level.

Daniel Levine:

Let's dig a little deeper there. This is not an estimate, but an actual count you undertook. If someone wanted you to name the 10,869 disorders, and I'm not going to have you do that, but you could list them. Walk me through the methodology and rigor for how you arrived at that number, what was included and what wasn't? I know there were some judgment calls involved here. Give me a sense of what those judgment calls were.

Seb Lefebvre:

As I mentioned earlier, we started with OMIM and Orphanet for a reason. Orphanet is orphan disease focused, and a lot of the rare disease are genetically focused, where OMIM plays another key role.

So, first and foremost, we had to clean up those data sets. There's a lot of entries that are inactive, obsolete, have been removed or have been moved because there's a better way or a better description of the disease, and they realized these two entries are the same, so they removed one—things like that. We cleaned up Orphanet, but we also did the same with OMIM. But with OMIM, we went a bit further. Because it's a genetic condition, it doesn't mean it's rare. So, we had to remove the non-disease, the non-rare, and we also removed the susceptibility to, from OMIM. All of this was done using the information that OMIM provides us, whether on the website or from our ability to download the dataset. But we also needed to manually review around 1,000 OMIM entries, to assess whether or not they were non-rare or non-disease. And when we say non-disease, we say things like trait. It could be that this variant or this gene modification causes a different color of your eyes or your hair. These are traits. By removing the non-rare and the non-disease from that 1000, we removed about 478 non-rare and non-disease from there. When we did this manual curation, we really focused on sensitivity. For

example, if we couldn't find anything around prevalence or the incidence about the condition, then we considered it to be rare.

Once the data set was cleaned up, we started by including all Orphanet disorders without subtypes, and the disorders with subtypes. And that gave us 6,282 conditions. We then went to OMIM and took only the OMIM conditions that had no overlap with Orphanet, that gave us 2,067 conditions. And then lastly, we looked at the OMIM entries that were a narrower match to an Orphanet entry. In this particular case, it would be considered as a subtype, and obviously, removed anything that was exact to Orphanet. We removed the parent of that, so for the Orphanet disorder and the OMIM narrow match, we didn't count the parent. We just counted the subtypes. And that gives us 2,520 condition. If you add these three numbers, you basically get to 10,869 conditions.

Daniel Levine:

Let's bring Wendy in here. As someone who heads patient advocacy at a leading rare disease drug developer, why does getting a more accurate number of rare disorders matter to patients and care caregivers?

Wendy Erler:

Thanks so much. And Seb, incredible work that you've been able to lead. As we've been talking about, we can acknowledge and all agree, the number of rare disease really is in flux. And fundamentally, without more accurate tracking and identification, it's patients who suffer and families who suffer. As scientists discover more diseases and refine their understanding of ones that we know now today, we continue to identify, but we really haven't updated the number.

And then, a related issue is that scientists and researchers don't always define or identify various diseases in the same way. So, there's a lot of terminology, a lot of registries, a lot of patient reported data, and each may have its own strategy for description.

Rare diseases often aren't included in standard clinical terminologies. And we know and fundamentally believe better counting of rare diseases will lead to better patient outcomes. Also, as somebody who works for a drug developer, I can state unequivocally, better understanding and counting of rare diseases will lead to more investment in research of these rare diseases, which then can hopefully lead to better treatments.

Daniel Levine:

Help me understand the drug developer's perspective on this. How does having a true picture of the number and impact of rare diseases inform the opportunity to develop new medicines?

Wendy Erler:

If we think about it and simplify it, because Seb shared a lot of complicated information that's really data driven and makes sense. As he mentioned, we know the vast majority of rare diseases have a genetic origin. About 85 or more percent of rare diseases are genetic. That means there's a lot of family information within that diagnosis and understanding that disease. And single gene disorders, those caused by a mutation or a defect in a single gene, we know those genes code for a normal cellular biological process. But something's gone wrong.

All of that is information. And with information, we drive the interest and commitment to research. And so, the more information we have and the more a drug developer and a pharmaceutical company can know about a specific disease, the more availability there are biomarkers and other pieces of

information, the better we can plan a research path and then seek investment in and commitment to delivering that research. So, it all ties together. But if we don't have that accurate number and way to describe rare diseases, we're still just throwing everything at the wall, instead of having a systematic path to being able to commit to research and development in a particular rare disease.

Daniel Levine:

One of the things that emerged from this paper was a map of the journey where disease communities should think about as they move from having an unrecognized disease to one that has treatments and even a cure. Can you explain that? How does this work inform learning about more rare diseases, and why does it matter?

Wendy Erler:

I think this is the part that's the most important. Because at the end of the day, we're talking about rare diseases, we're talking about numbers. But what we're really talking about is families, patients, and often, parents. The rare disease journey has been described as an odyssey. And that word alone really does bring into focus what we're talking about. And so often, families recognize that something is wrong. They know that there's something not quite right. It's usually with their child or loved one. Once they recognize that and they start to seek answers, this odyssey really begins. Next comes this process of trying to obtain a diagnosis. It points to the importance of being counted and having all of these diseases, as much as possible, continue to be recognized, because if we can shorten that time to diagnosis, we improve this odyssey or experience. And this is often self-led. It's parents, and family members, and people who care driving the process of obtaining the diagnosis by a navigating complex healthcare system. People have to self-advocate, they have to advocate for access to specialists, they have to question those people in white coats who say, "It's in your imagination" or, "It's this" when they know it's not that. It's a really emotional burden; often comes at huge personal financial expense and high emotional cost. And then, you said the word cure, that's a dream for most. But most of these rare diseases don't even have a treatment, much less a cure, and finding that treatment when none currently exists, is really a big part of this process. And this is where we have to take patient centricity from being words on a wall, to really recognize, patients and caregivers are the experts in this whole odyssey. They come to the table with accurate data, validated information, true disease experience. Really seeing this expertise as a piece of the puzzle, and then not having to reinvent the wheel as we seek to learn more, is really critical. So, I think that the journey part that patients and families go through can in a small way be aided by having a more accurate count and name to these rare diseases.

Daniel Levine:

Charlene, let me bring you in. RARE-X is a collaborative platform for global data sharing. Why did it undertake this study?

Charlene Son Rigby:

Thanks Danny. RARE-X has developed an open platform for collecting patient-reported data on rare diseases. And we're doing this across disorders. But beyond our technology, really at our roots, we're a patient advocacy organization. Our mission is to collect, structure, and share critical patient data at scale. Our intent is to dramatically accelerate understanding of diseases, and of course, therapy development. We're doing this in a way that enables patients to own their data. This gives them a real seat at the table for research. At the same time, we're trying to de-burden them, so that they can do this

data collection without having to become experts in data governance, research protocols, survey development, data standards, et cetera.

Going back to this point about this study. Ensuring that the true magnitude of rare diseases is understood and accepted is really critical. It's foundational to progressing work for rare disease patients and rare disease communities. That's why this study is called Be Counted. Rare disease groups must be counted to be recognized. Wendy did such an excellent job of describing all the social and psychological challenges. Seb also talked about the criticality of this as we move toward precision medicines. I want to point out that RARE-X supports data collection for undiagnosed patients. Our goal is to support their effort to get a diagnosis and really to get a name for their condition. This is deeply personal for me as we spent three years searching for a diagnosis for my daughter, who was finally diagnosed with a rare neurodevelopmental disorder. Getting a name for one's condition, and really getting recognition that one has a distinct and specific condition, is truly the first step on the road to therapies.

Daniel Levine:

As you think about the continuum that Wendy talked about, the movement of a disease from being unknown to identified, to diagnosable, to treatable, what role does data play and the type of data gathering and analysis RARE-X is enabling?

Charlene Son Rigby:

Characterizing diseases, which is identifying and understanding symptoms and understanding the progression of a condition, is really critical to advancing this continuum from unknown, to diagnosed, to treatable. There is a differentiation that we made around poorly defined versus diagnosable conditions in the study. On some level, diagnosable is a very minimal bar. To get from diagnosable to treatable requires comprehensive and fine grained characterization of a condition. Importantly, this data needs to be collected in a high quality and research ready way. What does this mean? At RARE-X, we've invested significantly in mapping to data standards like the Human Phenotype Ontology, like OMIM, like HL7, and other standards that will really facilitate research to generate insights out of this data, and to also facilitate aggregation of data across data sets. We've deployed a truly open data analysis platform. And this enables us to break down silos of data and really break through traditional challenges to data access. As a rare disease community, we have urgency to spur research and researchers forward. And one key way to do this is by making robust data easily yet appropriately available.

Daniel Levine:

The other thing that emerged was the essential role a small number of databases play in listing a condition and lifting it from obscurity, and the importance they can play in catalyzing research, and enabling diagnosis, and improving treatments. Seb, can you speak to that?

Seb Lefebvre:

Yes. As you look at how people are diagnosed today, the doctors and clinicians everyday use this process called differential diagnostic where you compare and contrast diseases that could be associated with the patient. If that information's not available, then no tools or website will be there, and the disease will not show up as an option. So, these databases are essentially providing that information, whether genetic or phenotypic, that is essential for these tools, but also the websites like OMIM and Orphanet, to support that process. So, I would say getting a disease's genetic and phenotypic information, any

disease databases, start with patient registries, for example, but eventually make your way to OMIM or Orphanet, which is key to supporting that differential diagnostic process that you see happening in a clinic every day.

Daniel Levine:

I'll throw this question out to everyone. But what steps should patient communities take to get their diseases on the map and drive research, and drug developers, and providers to act? I'd like to hear from all of you. Let's start with Seb.

Seb Lefebvre:

Yeah. Like I was just describing earlier, the information has to be found somewhere. As a patient community, I think the first step has to be to gather your data and get it somewhere. And the way to do that could be through patient registries, and partnering with researchers in your own disease area. But definitely get the genetic and the phenotypic information captured.

Once you've got that, then the ball starts rolling. And then publication occurs, and then that gets picked up by the OMIM and the Orphanet databases of this world, which then gets picked up by the tools, which then at the end of the day, helps the whole drug discovery and treatment process.

Daniel Levine:

Wendy, how about you?

Wendy Erler:

I think this is probably the crux of the most important question. It really starts with building that community. And when we think about the definition of community, it's bringing a group of people together that share common interests and goals. That's patients, that's researchers, that's advocacy groups. It's anybody that can come together and advance this goal.

For individuals, particularly those with ultra-rare diseases, there's so much responsibility that they take on to fight to be heard. My personal advice is to keep making those phone calls, ask the hard questions, fight for a seat at the table until you really do see that momentum moving.

I had the privilege of meeting a family a couple years ago who reached out and their daughter was diagnosed with a very ultra-rare condition. They were from Panama. My company was not researching in that area. Quite frankly, they showed up at our door and we met with them and we were able to do some work to help them use an algorithm to get an exact diagnosis and connect them to people who might potentially be interested in research. So, at the end of the day, some of this is all about connection and finding those resources. We're happy to help any way we can.

Daniel Levine:

Charlene, what are your thoughts?

Charlene Son Rigby:

I'm in absolute agreement with Seb and Wendy and just to build on Wendy's point around building community. Oftentimes the start for individuals and families is finding other people via social media. It's shocking to see how social media has changed the face of rare disease in the last five years. People can find other people based on a diagnosis, or based on similar symptoms, or clusters of symptoms. It's such a powerful way, since we have so few rare disease patients of any individual disorder in a specific

geography. It's an amazing way to just break through those geographic limitations, and really find people around the globe.

When we got my daughter's diagnosis, after feeling very, very much alone, it was so amazing to be able to find a Facebook group with other STXBP1 families. So, I highly recommend that as an important step of building your tribe. And then the other thing that I would say, we've talked a lot about contributing your data here and interfacing with researchers. One thing also that I've seen that's very exciting is the rise of citizen scientists, advocates who are really starting to engage in a very meaningful way with research and not waiting for other researchers to get interested, but really spurring forward, and frankly, sparking the research on their condition.

Daniel Levine:

Seb, we've talked about 80 percent plus rare diseases being either from genetic or suspected genetic causes. What does this work suggest about the importance of genetic sequencing, and particularly for patients who are not yet diagnosed seeking such test?

Seb Lefebvre:

The data speaks for itself. 61 percent of rare diseases, according to our analysis, have known genetics. That is a very important biomarker as part of your diagnostic odyssey. So, absolutely getting sequence is essential. Then there's 25 percent of those rare disease with suspected genetics. If we want to clearly define whether they're genetic causes or not and what gene's involved, and what's the inheritance pattern, we have to get those patients sequenced as well, in order to clearly determine the underlying genetic cause. So, as you say, 80 percent are rare genetic diseases so you cannot avoid sequencing. It's key.

Wendy Erler:

And I think I'd add that the path that most rare diseases follow take the rare disease from obscurity to a condition that is well understood. And by well understood that means researchable and potentially treatable, and then readily diagnosed. So, it all falls into line with the really important work of this paper is inclusion in the knowledge base of rare diseases is the first milestone to accelerate this path. We know and acknowledge this number will increase over time because there's so many new advances in diagnostics, biomarkers, genomics, proteomics. Science is moving fast. Rare diseases collectively represent a very big burden on the healthcare system and are a healthcare emergency. So, this work is really important.

Daniel Levine:

And Wendy, is there advice you could offer patient advocates, patient families, who want to see drug companies develop treatments for their particular diseases? What could they do to engage with drug developers and attract them to doing work in a specific condition?

Wendy Erler:

There's a lot of complexity to that question, because there's so much strategic initiative that goes into company decisions about which areas they develop drugs in, what their technology is, what their manufacturing capacity is, a whole lot of things. So, it's really complex. But from the patient perspective, I always say to patients and patient advocacy groups, "Reach out." Find the patient advocacy leaders in these organizations, or the head of R&D, whomever that leader is that will have the conversation, and whatever you can bring to the table, whether it's biomarker information. If a patient advocacy group has

started a registry, that's a robust suite of data. Anything that can help see how much about the condition is understood today. And then another opportunity is where's this likeness around one condition to another. So, if we've learned a lot about one rare disease, how does that impact and accelerate learning of another? Anything that's known about the genetics is really helpful. Then, it's about holding people accountable and saying, "What do you need from us to commit to research in this disease?" And getting the right voices and the right people at the table to listen.

Daniel Levine:

Charlene, what does this research say about the importance for rare disease patients to share their data?

Charlene Son Rigby:

We hear the term clinical trial readiness a lot now. So, working back from there, one of the key foundational aspects of developing clinical trial readiness is collecting enough data to characterize a disorder. As we were just talking about bringing that data into a registry and endpoints, which are important in clinical trials to measure the success of a potential therapy, can only really be developed once those phenotypes are well understood. So, with small and sometimes very small patients in any given disorder, every patient matters; every patient's individual symptoms, their experience, the progression of their condition, is really important to building that robust picture of a disorder.

From a registry perspective, the size of a registry may be used to estimate disease prevalence, as well as how activated a community is and also how willing a community is to participate in research so it's really paramount for rare disease patients to contribute their data, to speed understanding, research, and ultimately, therapy development.

Wendy Erler:

And Charlene, I would add, when we talk to rare disease patients and families, they want to contribute their data. They want all of this information to be shared to help others and I think there's also this tandem work of contributing to organizations like RARE-X, so these data are available in a federated nonprofit open source way, so that we can really accelerate research.

Daniel Levine:

The paper is Be Counted. You can find it on the RARE-X website under the Our Work tab. Sebastian Lefebvre, senior director of data science for Alexion, Wendy Erler, vice president of patient experience for Alexion, and Charlene Son Rigby, CEO of RARE-X, thank you all for your time today.

Wendy Erler:

Thank you everyone.

Charlene Son Rigby:

Thank you.

Seb Lefebvre:

Thank you.

Daniel Levine:

Thanks for listening. RARE-X is a collaborative platform for global data sharing and analysis to accelerate treatments for rare disease. RARE-X is adapting proven technologies and partnering with leading experts to create a federated data analysis platform, specifically designed by rare community leaders, to scale and to support the diverse and expanding needs of rare disease research, development, and care. To learn more about RARE-X, go to rare-x.org. This podcast is produced for RARE-X by the Levine Media Group. Music is courtesy of the Jonah Levine Collective.