A more accurate count of rare diseases and steps to getting counted

Kirk Lamoreaux[1], Sebastien Lefebvre, M.S.[1], Daniel S. Levine[1], Wendy Erler[1], Tom Hume[1]

[1]on behalf of RARE-X

Purpose

Develop an accurate estimate of rare diseases to raise awareness of the undercounted number among researchers, policymakers, and insurers, while empowering patient communities to get their diseases counted. This paper provides a transparent and reproducible approach to counting rare diseases.

Background

Governments, nonprofits, and industry organizations involved in rare disease research often state that there are 7,000 rare diseases. Thousands of conditions already included in rare disease knowledge bases (e.g., OMIM, Orphanet) are excluded when we repeat the 7,000 estimate. Unless a disease is included and described in a principal rare disease knowledge base, it is unlikely to be diagnosed, even by the best specialist. There is a path that most rare diseases follow that takes them from obscurity to a condition that is well understood, readily diagnosed, and treated.

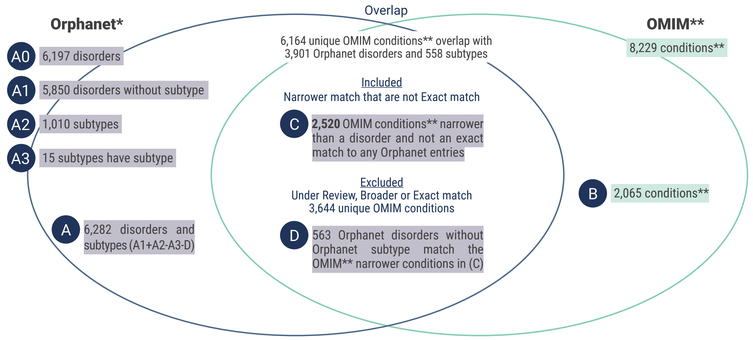
Methods

- Our methodology attempted to balance the inclusion of all rare diseases (sensitivity) with how we define a "disease" (precision) to minimize the risk of counting duplicates as a result of variations in disease names across knowledge bases.
- We leveraged a Rare Disease Map that integrates multiple public and licensed data sources to estimate the number of rare diseases. Orphanet and OMIM were the two primary sources for disease entries. In some instances, disease classifications were manually verified in GARD.
- For treatment information, we consulted DrugBank, the U.S. Food and Drug Administration's Office of Orphan Drug (OOD) website, and curated treatment information included in Genome to Treatment (GTRx). Our harmonization efforts leveraged Orphanet's mapping of Orphanet to OMIM entries as of Dec. 11, 2021.
- In our conservative base case, we counted all Orphanet disorders without ANY subtype and only the subtypes (or "children") of a disorder. We excluded groups of disorders, all "parent" disorders, as well as all subtypes with subtypes. We then added the OMIM narrow match set ("children" of an Orphanet disorder or subtype) and the OMIM-only set (OMIM entries without an Orphanet match).
- We applied additional filters to the remaining set to make sure no "parent" entry or duplicates (exact match) were included. We then characterized how well a condition, disease, or subtype is defined in Orphanet and OMIM based on its associated phenotypes relying on terms from the Human Phenotype Ontology (HPO). By analyzing the HPO terms included in the description of a disorder or subtypes, we have categorized all counted diseases as either "Poorly Defined" or "Diagnosable."

Methodology

Reconciling OMIM and Orphanet

Data Locked December 11, 2021



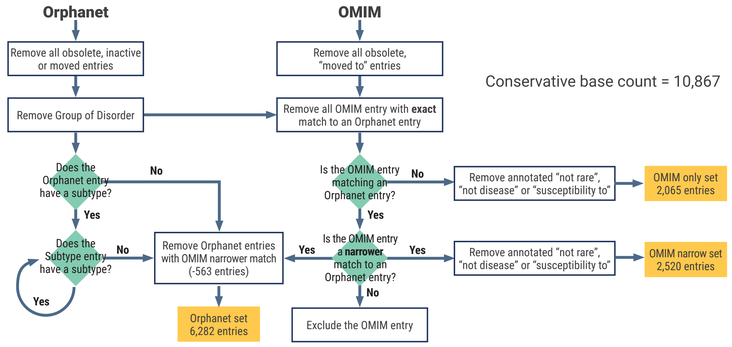
* removed "inactive", "obsolete", "moved to"

** "susceptibility to", "non disease", "moved to" (based on OMIM label), "non rare", "non disease"; 740 OMIM conditions have been manually annotated as "non rare", "non disease" or "susceptibility to" and excluded

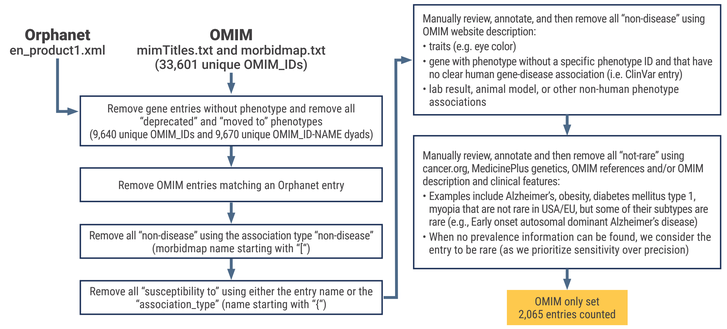
Conservative base case: A + B + C = **10,867**

Inclusive base case: A0 + A2 + B + C = **11,792**

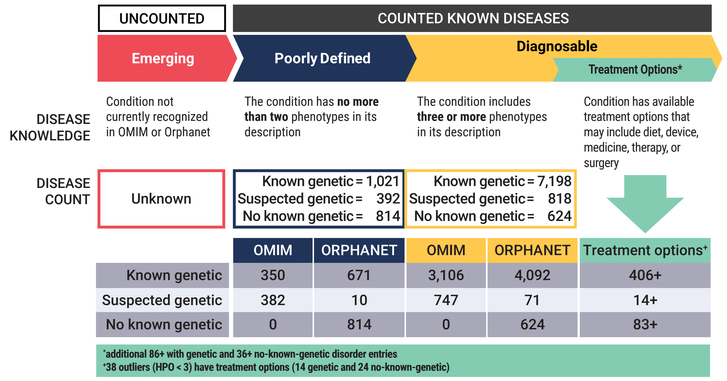
Methodology for Reconciling OMIM and Orphanet



Process for semi-automated curation of the "OMIM-only" data set



Breaking down of 10,867 counted rare diseases by associated phenotypes and genetic basis



	OMIM	ORPHANET	OMIM	ORPHANET	Treatment options*
Known genetic	350	671	3,106	4,092	406+
Suspected genetic	382	10	747	71	14+
No known genetic	0	814	0	624	83+

*additional 86+ with genetic and 36+ no-known-genetic disorder entries
*38 outliers (HPO < 3) have treatment options (14 genetic and 24 no-known-genetic)

Results & Findings

We arrive at our conservative base estimate of 10,867 conditions where 6,282 conditions are Orphanet entries, 2,065 conditions come from OMIM, and have no match to Orphanet, combined with 2,520 OMIM conditions that are considered a narrower match that might be reflective of a subtype. Our analysis found that 8,640 (80%) have disease descriptions at the parent or subtype level that included three or more phenotypes (labelled "diagnosable") and 2,227 (20%) rare diseases had less than three phenotypes (labelled "poorly defined"). Of the total count, 8,219 (76%) have a "known genetics", another 1,210 (11%) have "suspected genetics" and 1,438 (13%) are "non-genetic" diseases.

Conclusion

We believe there is inherent power in being counted, and we have provided a transparent, semi-automated and reproducible approach to counting rare diseases so that all stakeholders can confidently embrace the reality that there are approximately 11,000 rare diseases. This number is dynamic and will continue to grow, which is a testament to the success of researchers, patients, and the communities that advocate on their behalf. Inclusion in the knowledge bases of rare diseases is the first milestone on being counted.