Lawrence Berkeley National Laboratory

Title

How many rare diseases are there?

Permalink

Journal

Authors

Haendel, Melissa

Vasilevsky, Nicole

Unni, Deepak

et al.

Publication Date

DOI

How many rare diseases are there?

Melissa Haendel[1,*], **Nicole Vasilevsky**[1], **Deepak Unni**[2], **Cristian Bologa**[3], **Nomi Harris**[2], **Heidi Rehm**[4], **Ada Hamosh**[5], **Gareth Baynam**[6], **Tudor Groza**[7], **Julie McMurry**[8], **Hugh Dawkins**[9], **Ana Rath**[10], **Courtney Thaxon**[11], **Giovanni Bocci**[3], **Marcin P. Joachimiak**[2], **Sebastian Köhler**[12], **Peter N. Robinson**[13], **Chris Mungall**[2], **Tudor I. Oprea**[3,*]

[1]Oregon Clinical & Translational Science Institute, Oregon Health & Science University, Portland, OR, USA.

[2]Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

[3]Department of Internal Medicine and UNM Comprehensive Cancer Center, UNM School of Medicine, Albuquerque, NM, USA.

[4]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.

[5]McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA.

[6]Western Australian Register of Developmental Anomalies and Genetic Services of Western Australia, King Edward Memorial Hospital, WA Department of Health, Perth, Australia.

[7]Pryzm Health, Perth, Australia.

[8]College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA.

[9]HBF Health Limited, Perth, Australia.

[10]INSERM, US14-Orphanet, Paris, France.

[11]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA.

[12]Charité Centrum für Therapieforschung, Charité— Universitätsmedizin Berlin Corporate Member of Freie Universität Berlin, Berlin, Germany.

[13]Jackson Laboratory for Genomic Medicine, Farmington, CT, USA.

Abstract

*melissa@tislab.org; TOprea@salud.unm.edu.

A lack of robust knowledge of the number of rare diseases and the number of people affected by them limits the development of approaches to ameliorate the substantial cumulative burden of rare diseases. Here, we call for coordinated efforts to more precisely define rare diseases.

The phenotypic, genetic and environmental characteristics that define a given disease are often established in different demographics, regions or contexts. Multiple observations in different human subpopulations corroborate disease definitions, supporting standardized diagnostic capabilities and therapeutic approaches. However, the probability of such corroboration for rare diseases is substantially lower. The lack of agreement and imperfections in current rare disease definitions affect patient resources, diagnosis and treatment approaches, and research on rare disease mechanisms and potential therapies.

As patients with rare diseases are estimated to constitute as much as 10% of the population (see Related links below), there is increasing societal pressure to develop diagnostics and therapies for rare diseases. Consequently, there is an urgent need to more precisely define rare diseases and to improve access to accurate information on such diseases in order to realize the goals of precision medicine for the many affected people. However, at present, it is not even possible to accurately say how many rare diseases there are.

How rare is rare?

The most common definitions for what constitutes a rare disease are the eligibility criteria used by drug regulatory agencies for financial and regulatory incentives to develop rare disease therapies (known as orphan drugs). Such definitions differ between different regions. For example, a rare disease in the United States is defined by the 1983 Orphan Drug Act as a condition that affects fewer than 200,000 people, whereas the analogous legislation introduced in the European Union in 2000 considers a disease to be rare when it affects fewer than 1 in 2,000 people. Furthermore, some diseases are 'rare' in some demo- graphics or regions, but not in others. For example, Tay–Sachs disease is rare in the general population, but has a carrier frequency of 1/25 in Ashkenazi Jews, and tuberculosis is rare in the United States, but is one of the top 10 causes of death worldwide according to the World Health Organization (see Related links). Clearly, rarity is contextual.

What constitutes a rare disease?

Rare Mendelian diseases are caused by specific genetic variants, in contrast to most common diseases, which are associated with numerous small-effect genetic variants and for which the

influence of environmental factors is often more important. Other types of rare disease such as rare infectious diseases and rare cancers have their own characteristics and complexities. All rare diseases, however, share similar clinical challenges, as most clinicians are unlikely to have experience with recognizing or managing the vast majority of rare diseases. Diagnosis is often delayed or wrong, and optimal clinical management is seldom achieved.

As genome sequencing becomes routine, it is expected that doctors seeking to diagnose patients who may have a rare disease will begin by sequencing a patient's genome and then attempt to identify associated diseases based on the patient's phenotypic features. However, even with the current limited knowledge of the genetic basis of many rare diseases, it is known that different pathogenic variants in the same gene may have different consequences, which is often not adequately recorded. For example, different pathogenic variants in the PIK3CA gene are associated with several different rare diseases, including Klippel–Trenaunay syndrome, the megalencephaly–capillary malformation syndrome and several cancers and malformation syndromes (see Related links). Although involving variants in the same gene, these should be considered distinct diseases because their presentation and, more importantly, their treatment are different.

In order to more effectively diagnose and care for patients with rare diseases, we first need to agree on the definition of the unique combination of genetic, phenotypic and environmental attributes associated with each rare disease, including how we define disease incidence[1]. Currently, there is a plethora of overlapping terminologies, models and metadata for identifying and classifying rare diseases[2], in juxtaposition with increasing data harmonization for proteins and genes that could represent therapeutic targets[3]. Although disagreement about the definitions of certain diseases seems inevitable, there is a pressing need for a computational representation of disease that can integrate multiple resources to promote comprehensive information retrieval from all relevant sources, as well as consistent diagnostic and care recommendations.

How many rare diseases are there?

Regulators, scientists, clinicians and patient advocacy groups often cite ~7,000 as the number of rare diseases, or between 5,000 and 8,000 depending on the source (see Related links below). Why do estimates of the number of rare diseases vary so widely? One reason is the afore- mentioned lack of consistency in defining discrete dis- ease entities and their incidence in different countries or demographics. Another reason is computational differences and imperfections in the current terminologies. Some terminologies do not include chromosomal disorders, or other structural variations such as inversions, while others do not include rare diseases with environmental causes, such as toxin exposure. Efforts to count the number of rare diseases do not delineate the inclusion/exclusion criteria with the computational attributes in mind, nor do they state whether the count includes more general disease categories or not, or even how these were defined.

Recently, a large number of terminological resources have come together to harmonize disease definitions in the Monarch Disease Ontology[4] (Mondo; see Related links). While this consensus process is still ongoing, we currently estimate the number of rare diseases to

be more than 10,000 (see Supplementary Box 1 for details of the analysis). This number was obtained by taking advantage of the hierarchical structure of existing disease terminologies. In the absence of a globally accepted definition for rare diseases, an initial pragmatic approach is to count the most specific terms (that is, 'leaf terms') in the disease hierarchies, while excluding higher level terms. For example, 'tetralogy of Fallot', a specific rare disease term in Mondo, is counted, but not its parent term, 'congenital heart disease'. When such information from the major knowledge sources on rare diseases — including Orphanet, OMIM, GARD, DOID and NCI Thesaurus — is combined algorithmically and curated within Mondo, a total of 10,393 rare disease 'leaf terms' can be identified. The majority, 6,370 rare diseases, are present in three or more resources, whereas 4,023 are unique to one source (see Supplementary Box 1). This preliminary analysis suggests that there could be a substantially higher number of rare diseases than typi- cally assumed at present, with obvious implications for diagnostics, drug discovery and treatment5. However, it should be emphasized that much more rigorous analysis is needed to establish if this is indeed a more accurate estimate.

A call to action

It is clear we must join forces globally and across resources, clinically and computationally, to collect, consolidate and curate the most robust and up-to-date knowledge on rare diseases. For example, OMIM adds more than 200 new disorders to its database annually, and it is essential that such knowledge sources are regularly harmonized.

We call here for funding and regulatory agencies, patient advocacy groups and all other organizations in the rare diseases field to support a coordinated effort to define, in a geopolitically neutral manner, what constitutes a discrete rare disease. A globally consistent set of criteria that enables rare diseases to be precisely defined will enable efforts to count rare diseases based on existing knowledge, as well as harness the rapid growth in genomic knowledge to elucidate the molecular basis of currently unknown or poorly characterized rare dis- eases. This will provide a foundation for more effective diagnosis and care of patients with rare diseases, and the development of new therapeutic approaches. Ultimately, if knowledge on rare diseases is not collected and curated more effectively, many patients with rare diseases will remain underserved or neglected by healthcare systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Haendel MA, Chute CG & Robinson PN Classification, ontology, and precision medicine. N. Engl. J. Med 379, 1452–1462 (2018). [PubMed: 30304648]

2. Haendel MA et al. A census of disease ontologies. Annu. Rev. Biomed. Data Sci 1, 305–331 (2018).

3. Oprea TI et al. Unexplored therapeutic opportunities in the human genome. Nat. Rev. Drug Discov 17, 317–332 (2018). [PubMed: 29472638]

4. Mungall CJ et al. k-BOOM: a Bayesian approach to ontology structure inference, with applications in disease ontology construction. bioRxiv 048843; 10.1101/048843 (2019).

5. Tambuyzer E et al. Therapies for rare diseases: therapeutic modalities, progress and challenges ahead. Nat. Rev. Drug Discov (in the press).