If we cannot count rare disease patients, they will not count

Melissa Haendel, PhD, FACMI
1/10 americans (400M globally) affected

Each patient’s characteristics are akin to the variation in zebra stripes.

“When you hear hoof beats, think horses, not zebras.”

“There are ~7000 rare diseases”

- per Orphan Drug Act (1983);

This number is demonstrably wrong.

The case for open science: rare diseases
Why the number of rare diseases is hard to determine (and is not 7000!)

We don’t have the same criteria for “rare” around the world:
- 1983: From the Orphan Drug Act: A rare disease affects fewer than 200,000 people
- 2000: European Union considers a disease to be rare when it affects fewer than 1 in 2,000 people.

We add new diseases all the time, but don’t update the number:
- New rare diseases are discovered every week by organizations such as the Undiagnosed Disease Network
- The literature and public databases abound with new weekly entries
- N-of-1s are matched, defining new diseases

We don’t define or IDENTIFY diseases in the same way
- Dozens of terminologies and disease registries exist, each with their own identifier systems or lack thereof
- Rare diseases are often not included in standard clinical terminologies (such as ICD)
- Fundamentally, the definition of a rare disease and how to model it computationally has remained more an art than a science, preventing interoperability
Jessica (age 4) has a rare condition which causes epilepsy, affects her movement and developmental delay. Standard genetics tests negative.

- De novo deletion in *SLC2A1* identified as the cause of her Glut 1 deficiency syndrome
- Now being successfully treated with a ketogenic, low-carb diet
- Low risk for future pregnancies

It can take patients 4-7 yrs to get a diagnosis
Every person is an n-of-one disease (this is the promise of precision medicine)

The question is what are meaningfully groupings of patients?

signs and symptoms, demographics, exposure, diet, traits, etc.
We need to leverage ALL biological knowledge about the relationships.
Turning information into meaning

Flat List

- e.g., Problem lists, drugs

Taxonomy

- e.g., ICD

Graph

We are mostly here

We need to be here
The challenges start with the basics: Phenotyping
Patients and families should be empowered to work as partners...

The time is now!

- Provider time constraints
- Care fragmentation
- Knowledge democratization
- Tools and standards emerging to help
Answer:
Give patients the ability to describe their conditions in their own words while leveraging formal vocabularies used by clinicians.

Human Phenotype Ontology (HPO)

- Phenotyping terminology
  >14,500 terms

- Computational disease models
  >190,000 disease-phenotype annotations (associations)

- Widely adopted in rare disease genomic diagnostic tools
  100,000 Genomes Project, SOLVE-RD, NIH-UDP, etc.
When describing a patient, how many terms can you take away and still diagnose the disease?
An example comparison of simulated profiles relative to gold standard

Legend

- Perfect Match
- Fuzzy Match
- No Match

doi:10.6084/m9.figshare.5513356.v2
When any profile is further constrained by deleted terms or less specific terms general, the diagnostic power reduces as expected.

What is the diagnostic power of the layperson HPO?
Phenotype-assisted diagnosis in action

Not same variant, but same disease and gene, KMT2A.

DOI: 10.1126/scitranslmed.3009262

Legend

- Perfect Match
- Fuzzy Match
- No Match

For diagnosis to work like this, the reference knowledge for each disease needs to be systematically queryable.
But ... 
Assembling a comprehensive source of disease definitions (and annotations) is hard.
Different communities annotate different relationships, at different levels of granularity and using different vocabularies.
The knowledge about diseases come from different sources / communities, often siloed by the type of disease. We needed:

- Disease concepts spanning multiple categories
- A systematic way of relating these concepts and resolving inconsistencies, especially where 1:1 equivalence not possible

mondo.monarchinitiative.org
Defining diseases: Lumping and splitting

- Distinct molecular mechanisms?
- Reputable assertion of difference?
- Distinct clinical management?
- Distinct phenotypic profiles?

LUMP

SPLIT

bit.ly/lump-split
Are they they same disease?

- Different labels
- Different parents
- Different children
- Different synonyms
- Different text definitions

Are they equivalent?

https://github.com/monarch-initiative/mondo/issues/61
Evidence-based, curated merging of equivalent disease concepts

kBOOM
Bayesian OWL Ontology Merging
Logical + Probabilistic Inference

evaluate

Iterative curator-assisted equivalence inference

curate
feedback

Curated Equivalence Relations

OMIM
Orphanet
NCIT
GARD
EFO
DO
DC
MESH
MEDIC
...

doi.org/10.1101/048843
Aligning disease knowledge across sources and tracking provenance:

Mondo concept for adult Refsum disease (MONDO:0009958)
Mondo

A curated, evidence-based merging to harmonize diseases and phenotypes across sources

mondo.monarchinitiative.org
Hierarchical classification of adult Refsum disease
If rare diseases are not counted, rare disease patients will not count

Many diseases are in only one source

Just 5 sources comprise 10,443 unique rare disease concepts (prior estimates ~7,500)

Only 333 shared disease concepts in all five sources

How many undiagnosed patients are there in your hospital?
From EHR data, we can estimate rare disease prevalence (for those with ICD codes) in a single system. Results often at odds with the published estimates.

<table>
<thead>
<tr>
<th>Rare Disease</th>
<th>Patient Count</th>
<th>OHSU Patient Prevalence (1,039,213 total patient count)</th>
<th>Published estimate of population prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystic Fibrosis</td>
<td>567</td>
<td>0.0151%</td>
<td>0.0090%</td>
</tr>
<tr>
<td>Sickle Cell Disease</td>
<td>259</td>
<td>0.0250%</td>
<td>0.0308%</td>
</tr>
<tr>
<td>Muscular Dystrophy</td>
<td>709</td>
<td>0.0680%</td>
<td>0.0769%</td>
</tr>
<tr>
<td>Lennox Gastaut Syndrome</td>
<td>503</td>
<td>0.0480%</td>
<td>0.0001%</td>
</tr>
<tr>
<td>Urea Cycle Disorder</td>
<td>93</td>
<td>0.0090%</td>
<td>0.0029%</td>
</tr>
<tr>
<td>Takayasu’s Arteritis</td>
<td>24</td>
<td>0.0020%</td>
<td>0.0002%</td>
</tr>
<tr>
<td>Pheochromocytoma</td>
<td>100</td>
<td>0.0100%</td>
<td>0.0005%</td>
</tr>
<tr>
<td>Hereditary Hemorrhagic Telangiectasia</td>
<td>341</td>
<td>0.0330%</td>
<td>0.0200%</td>
</tr>
<tr>
<td>Osteogenesis Imperfecta</td>
<td>122</td>
<td>0.0120%</td>
<td>0.0050%</td>
</tr>
<tr>
<td>Eosinophilic Esophagitis</td>
<td>354</td>
<td>0.0340%</td>
<td>0.0500%</td>
</tr>
<tr>
<td>Batten Disease</td>
<td>145</td>
<td>0.0140%</td>
<td>0.0030%</td>
</tr>
<tr>
<td>Focal and Segmental Glomerulosclerosis</td>
<td>153</td>
<td>0.0150%</td>
<td>0.0007%</td>
</tr>
<tr>
<td>Mitochondrial Neurogastrointestinal Encephalopathy</td>
<td>66</td>
<td>0.0060%</td>
<td>0.0010%</td>
</tr>
</tbody>
</table>
Rare disease prevalence varies across healthcare systems & public sources.

Differences in prevalence across healthcare systems are in part due to rarity and diagnostics, but also due to differences in coding.
Age at First Diagnosis Varies by Rare Disease

Large spread of initial diagnosis is due to both challenges in diagnosis, but also in obtaining correct information from the EHR.
Each patient is different: what are the patterns?
Patient diagnostic journey maps reveal patterns & differences across health care systems and diseases.
What about all the Rare Diseases without an ICD code?
We must get rare disease definitions into the EHR, but we have to start where we are.

It’s a Mad Max Sitch.
Extra
Plain language synonyms for patients to use

4887 of 13823 HPO terms have lay synonyms
The case for open science: rare diseases


Author Notes

Published: 11 September 2020

Abstract

The premise of Open Science is that research and medical management progress faster if data and knowledge are openly shared. The value of Open Science is nowhere more important and appreciated than in the rare disease (RD) community. Research into RDs has been limited by insufficient
Combining genomic and phenomic data improves variant prioritization for diagnosis

github.com/exomiser

doi: 10.1038/gim.2015.137
Evaluation of Exomiser in 100K genomes project: Made possible by open science

The 100,000 Genomes Pilot on Rare-Disease Diagnosis

U.K. Patients with Rare Diseases and No Diagnosis — Preliminary Report

<table>
<thead>
<tr>
<th>Diagnostic yield</th>
<th>25% of probands received a genetic diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic pipeline</td>
<td>86% of diagnoses were identified through automated pipeline</td>
</tr>
<tr>
<td>Novel discoveries</td>
<td>3 new disease genes discovered</td>
</tr>
<tr>
<td></td>
<td>19 new disease–gene associations identified</td>
</tr>
</tbody>
</table>

25% of genetic diagnoses had immediate ramifications for clinical decision making.

November 11, 2021

Mondo community paper
with ClinGen OMIM, and
the rest of the village!

https://www.medrxiv.org/content/10.1101/2022.04.13.22273750v3
Mondo accommodates gene-based names AND groups terms based on gene-based etiology.

For example, GTP cyclohydrolase I deficiency

- GTP cyclohydrolase I deficiency
- dystonia 5
- GTP cyclohydrolase I deficiency with hyperphenylalaninemia

ClinGen curation target

Mondo terms mapped to OMIM

# 128230
DYSTONIA, DOPA-RESPONSIVE; DRD

# 233910
HYPERPHENYLALANINEMIA, BH4-DEFICIENT, B; HPABH4B

Alternative titles; symbols
DYSTONIA 5; DYT5
DYSTONIA, PROGRESSIVE, WITH DIABETIC VARIATION

mondo.monarchinitiative.org
60% of diseases are covered with HPO plain language subset terms.

Actual mileage (for a given disease) may vary depending on its layperson-coverage of the corresponding phenotypes.

How diagnostically useful are the HPO terms generated by patients?

Survey Mapping | Layperson HPO | Clinical HPO >14,500 terms
---|---|---
Good | Better | Best

Phenotype Profile Granularity
A Census of Disease Ontologies

Annual Review of Biomedical Data Science
Vol. 1:305-331 (Volume publication date July 2018)
First published as a Review in Advance on May 9, 2018
https://doi.org/10.1146/annurev-biodatasci-080917-013459

Melissa A. Haendel,1,2 Julie A. McMurry,1 Rose Relevo,1 Christopher J. Mungall,3 Peter N. Robinson,4 and Christopher G. Chute5

1Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon 97239, USA; email: haendel@ohsu.edu
2Linus Pauling Institute, Oregon State University, Corvallis, Oregon 97331, USA
3Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
4The Jackson Laboratory, Farmington, Connecticut 06032, USA
5School of Medicine, School of Public Health, and School of Nursing, Johns Hopkins University, Baltimore, Maryland 21205, USA
Total costs incurred at a organization are dependent on the size of the patient population. While total costs for an individual rare disease cohort may not exceed an age-matched control cohort, collectively the rare disease cohort costs are greater than controls.
Rare diseases patients have increased healthcare utilizations compared to control patients. While costs are not available in N3C, the healthcare utilizations (encounters, medications, procedures, etc.) are available.
Rare Diseases collectively are an unrecognized public health crisis.
Under-utilized data → Loss of discriminatory power

Let’s change this
The IDeaS initiative: pilot study to assess the impact of rare diseases on patients and healthcare systems

Ainslie Tisdale¹, Christine M. Cutillo², Ramaa Nathan³, Pierantonio Russo³, Bryan Laraway⁵, Melissa Haendel⁶, Douglas Nowak⁷, Cindy Hasche⁷, Chun-Hung Chan⁸, Emily Griese⁸,⁹, Hugh Dawkins¹⁰, Oodaye Shukla³, David A. Pearce⁷,⁸,¹¹, Joni L. Rutter² and Anne R. Pariser*