UNIVERSITY OF LEEDS   University of Sheffield   UNIVERSITY of York

Original research

How do clinician and parent-reported data differ? An analysis of similarity and difference in the datasets from a cross-syndrome genetics cohort study (GenROC)

Karen Jaqueline Low [iD] ,[1,2] Huw Day,[3] Mevmi Lasanya Kodippuli Thanthilla [iD] ,[4] GenROC consortium,[5] Charlotte Davis,[4] Helen V Firth,[6,7] Caroline Wright[8]

ABSTRACT

Background Parent/patient-reported (PRD) datasets provide ready access to phenotypic data for monogenic neurodevelopmental disorders, yet their concordance with clinical data is unclear.

Methods In the GenROC study, 547 children (mean age 7.6 years, balanced sex ratio) had parallel parent-reported web questionnaires and clinician-reported (CRD) Human Phenotype Ontology proformas. We compared the two sources per participant by system, gene and gene group and overall for quantity, detail and similarity.

Results 547 probands were analysed ranging in age from infancy to 16 years (mean 7.6) with similar gender distribution. PRD provided more terms for dental, gastroenterology, immunology and respiratory systems and for vision (p<0.001 for all) and to a lesser degree for cardiac (p=0.0012). CRD provides more detail than PRD for most gene subgroups, combined systems and for neurology (p<0.001). Similarity scores were low overall per participant (mean 0.38 for combined). Similarity scores were highest for cardiac (mean 0.74) and lowest for Ear/Nose/Throat(ENT) (mean 0.34). There was minimal difference in similarity scores across gene groups or between the top 10 genes—scaffold adaptor gene groups had the highest (mean 0.43) as did *STXBP1* (mean 0.5) and *CACNA1A* (0.49). CRD is more similar to published syndrome phenotypes for syndromic genes.

Conclusions Parents reported more common childhood phenotypes, such as asthma and dental issues, while clinicians provided clinical phenotype descriptors, such as brain morphology and seizure semiology. It is important to understand the differences when designing studies and using datasets to appreciate their strengths and limitations.

INTRODUCTION

Data are limited regarding phenotypes and natural history in rare monogenic neurodevelopmental disorders (NDDs).[1 2] This data gap has led to many parent/patient-led foundations, as well as clinicians and researchers setting up parent/patient-reported (PRD) natural history studies, some of which are run independently and many of which are hosted within wider platforms such as RareX and Simon's searchlight.[3 4] Due to the obstacles and barriers to rare disease

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Parent-reported data are increasingly used in rare disease research due to their accessibility and breadth. Previous studies have shown that such data can be consistent with published literature, particularly in syndromic conditions. However, direct comparisons between parent-reported and clinician-reported data at the individual level have been limited, leaving a gap in understanding the reliability and granularity of these data sources.

WHAT THIS STUDY ADDS

⇒ This study provides the first large-scale, individual-level comparison of parent-reported and clinician-reported phenotypic data across a cross-syndrome cohort. It demonstrates that while both sources contribute similar quantities of data, they differ in content and detail. Parents tend to report common childhood and lived experience phenotypes, whereas clinicians provide more specific clinical descriptors. The study also shows that clinician data are more consistent with published syndrome phenotypes, especially in syndromic genes.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These findings highlight the complementary nature of parent and clinician data in rare disease research. Future studies and registries should consider integrating both sources to enhance phenotypic richness and accuracy. Policymakers and researchers designing data collection tools or machine learning applications should account for the strengths and limitations of each data type, ensuring that lived experience data are not overlooked in phenotype descriptions.

research, these studies provide important accessible data for researchers and clinicians.[5]

A scoping review identified nine studies[6–14] in monogenic NDDs which had included any comparisons of data provided from parents, clinicians and other sources. One of these studies has subsequently

BMJ Group

been cited by five publications regarding *SLC6A1* demonstrating the potential reach and impact on the literature of parent-reported data. Three further studies report on GenIDA collated parent-reported data in Koolen de Vries syndrome, *DDX3X* and Kleefstra syndrome and briefly report on data consistency with published phenotypes.[12–14] GenIDA is an international online data collection platform which aims to improve the understanding and management of genetic intellectual disability and autism disorders based on the information provided by caregivers and uses 46 in-depth multilevel questions.[15] Another platform, RARE-X, uses patient-reported outcome measures (PROMS).[16] Some rare disease groups have set up their own platforms and generated their own set of questions.[7 15 16] Only one of these nine studies[9] included any direct comparison on a patient level of parent and clinician-reported (CRD) data (table 1).

Given the drive towards parent-reported online survey data as the primary phenotype and natural history data source for NDDs, we sought to investigate the consistency, similarity and differences in granularity of data provided by parents and clinicians on an individual patient level. We used a cross-syndrome cohort in order to better inform understanding of and decision-making regarding data sources and their applications in NDDs in the future.

METHODS

Data were derived from the GenROC study, a UK-based cross-syndrome cohort study of children with NDDs under 16 years of age. All participants had a confirmed pathogenic or likely pathogenic change in a single gene identified through clinical testing or through another research study. The full protocol for the study is published elsewhere.[17] The GenROC study received Research Ethics Committee approval (22/EM/0274) on 15 December 2022 and Health Research Authority approval on 9 February 2023 (IRAS321040).

Parents were asked to complete an online web-based survey in Research Electronic Data Capture (REDCap) electronic data capture tools hosted at the University of Bristol. REDCap is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture; (2) audit trails for tracking data manipulation and export procedures; (3) automated export procedures for seamless data downloads to common statistical packages; and (4) procedures for data integration and interoperability with external sources.[18 19] The survey included a broad variety of questions regarding their child's development, growth and medical features. Some questions were checkbox choices (including sleep, feeding, behaviour, learning and education), but for each physical system, parents were asked to complete a free-text box. These free text boxes mirrored the same set of boxes that clinicians were asked to complete. They were given the option of leaving the box blank if not relevant for their child. Parent free-text responses for the cohort were then separated by clinical system to avoid later inferences from the curator across the systems. Three non-specialist curators (two medical students and one junior doctor) then reviewed the free-text data and curated them into Human Phenotype Ontology (HPO) terms as per the April 2024 version of hpo.jax.org. Curators were provided with free text responses by the system for participants, for example, working through just the cardiac responses at one time. This was done to limit cross-system inferences that could be made by the curators. A proportion of all the codes for each system was double coded by the senior investigator, a clinical geneticist, to check consistency across coding. For the purposes

Table 1 Summary of parent-reported data used in previous studies which compared parent data with the literature or clinical data. Only one, highlighted in bold, included any direct comparisons on an individual level

Study	Condition	Data collection source
Engwerda et al[8]	Proximal 6q deletions	Online 'chromosome 6 Questionnaire': parents recruited through social media group. Comparison made with case reports in the literature.
Engwerda et al[9]	Chromosome 6 aberrations	Comparison of 20 individuals whose parents completed the 'Chromosome 6 questionnaire' with the children' medical records. Also compared 34 individuals who took part in the questionnaire with 39 comparable aberrations reported in the literature.
Engwerda et al[10]	Terminal 6 q deletions	Online 'chromosome 6 Questionnaire': parents recruited through social media group. Comparison made with case reports in the literature.
Bain et al[11]	SLC6A1	Compared phenotypes reported in a 'provider reported' dataset comprising cases in the literature and in a database with 43 individuals who provided data via Simon's Searchlight.
Colin et al[12]	Koolen-de-Vries Syndrome	Caregiver-reported data for 237 individuals via GenIDA platform. Comparison made with literature descriptions.
Rualt et al[13]	DDX3X	Reported and compared two 'series' of patients. Series 1 (physician reported (48 individuals)) and series 2 (parent reported in GenIDA (44 individuals)).
Zdolsek et al[14]	Kleefstra syndrome	Parent-reported data on 172 individuals in GenIDA. Discussed findings in the context of the previous literature.
Vasireddi et al[6]	Kleefstra syndrome	Evaluated two series of patients for evidence of atrial tachyarrhythmias: series 1 (50 clinically reported patients at Radboud University) and series 2 (163 parent reported patients in GenIDA).
Low et al[7]	KBG syndrome	Parent and patient reported data from 91 adults with KBG syndrome collected and collated then compared with previous findings from clinically reported cohort of 36 adults.

of these analyses, parent-reported data refers to curated terms derived from parent free text responses.

For each participant in GenROC, their responsible clinical geneticist (or delegate) was invited to complete a clinical proforma via a web-based survey. This included genomic variant data alongside growth, prenatal and phenotype terms grouped by the same body systems as in the parent survey. Clinicians were requested to provide HPO terms but were given the option of free text response. Where free text response was provided, this was curated into HPO terms by expert curation (a clinical geneticist). These curated data were then processed in Python using the pyhop.ontology package (https://pyhpo.readthedocs.io/en/stable/). All the functions used (but none of the notebooks) are available (https://github.com/JGIBristol/GenROC_Public). HPO and variants are shared in DECIPHER.[20]

HPO terms from the clinician and parent were then analysed on an individual participant level using the following approaches. Probands where data were only collected from one source were excluded from this analysis.

For the purposes of the analyses below, we have grouped genes into gene group categories. The categories were defined using the following process. The gene list was first uploaded into PantherDB[21 22] which provides groupings of genes based on molecular function. After this initial step, unclassified genes were manually reviewed with gene categories. Categories were expanded to include further genes based on molecular and clinical characteristics. (See online supplemental table 1 for list of genes and curated categories).

The following scores were generated for each proband for PRD and CRD.

Quantity score

The number of HPO terms provided (high number=many HPO terms provided). We compared the CRD and PRD quantity result scores and counted the instances of 'Parent more', 'Doctor more' and 'Equal quantity" for System, gene category and top 10 genes. For all three, we calculated p values for 'Doctor more' and 'Parent more' using McNemar's χ^2 test in R. Given the multiple comparisons, raw p values were adjusted by the Benjamini–Hochberg false-discovery-rate procedure (FDR).

Detail score

A quantitative score reflecting distance in detail in the HPO branches between all possible pairs (one from PRD and one from CRD) of HPO codes for a given participant. For each pair of codes, we assess if one code is an ancestor of another code in the HPO tree. If one is an ancestor of the other, then the detail score increases for the source providing those data (PRD or CRD). The amount the score increases scales with the level of precision, for example, if PRD says 'Mild myopia HP:0025573' and CRD said 'Myopia HP:0000545', then PRD would get a detail score increase of one. However, if the CRD had instead said 'Abnormality of refraction HP:0000539' which is the direct ancestor of Myopia, then the PRD detail score would increase by two. If not, the detail score does not change. If, however, CRD has said 'Nystagmus: HP:0000639' and PRD said 'Mild myopia HP:0025573', neither would score as these terms are unrelated. We compared the CRD and PRD detail results scores for 'Doctor more' and 'Parent more' for System, gene category and top 10 genes. P values were calculated using McNemar's χ^2 test in R.[23] Given the multiple comparisons, raw p values were adjusted by the Benjamini–Hochberg FDR.[24]

Calculated semantic similarity scores: We used the similarity score system[25] for all pairs to assess similarity score by system and overall per participant. Pairwise semantic similarity scores ranging from 0 to 1 were calculated for gene pairs based on their associated HPO terms using the ontologySimilarity package in R (v4.2.1) (https://cran.r-project.org/web/packages/ontology-Similarity/vignettes/ontology Similarity-introduction.html). This method applies Lin's information-theoretic approach to quantify term similarity within the HPO hierarchy, where a score of 1 indicates identical phenotypic annotations.[9 17] A similarity score of 1 indicates an identical response. We manually reviewed 20 probands with the lowest similarity scores to assess for any common themes.

For the top 10 most recruited genes in GenROC, we compared the PRD, CRD and combined HPO terms per participant with the associated HPO terms in G2P for each respective gene (https://www.ebi.ac.uk/gene2phenotype/) and calculated similarity scores. We first applied a Welch one-way analysis of variance[26] in R to test global differences without assuming equal variances. Significant results were followed by all pairwise Welch t-tests, with Holm's sequential adjustment.[27]

Phenotype frequency by data source: we reviewed the numbers of HPO codes from PRD and CRD by system over the whole cohort. We reviewed which codes were most frequent and common to both sources and which were unique.

Finally, we manually reviewed the reporting of one common and important phenotype to determine reporting consistency and assess for differences. We selected epilepsy as this is frequent in the cohort and a phenotype that most parents would be aware of. For this analysis, we searched all HPO terms in PRD and CRD for HP:0001250 or a child code. This outputted a present/absent variable for each proband. For all 'absent' cases, an expert clinician also reviewed the HPO terms to check that none had been missed that were consistent with epilepsy. For all 'present' cases, these data were then stratified into PRD reported; CRD reported; and 'both' and manually by seizure classification where these data were available. A comparison was made across the three categories to assess for any differences.

RESULTS

547 participants had been recruited to the GenROC study at the time of this analysis, of which 477 participants had PRD and CRD. The amount of missing data varied by system as for some systems, one may have provided data but not the other and we elected to exclude these from systems analyses in order to allow for direct comparisons. For the 'combined' systems analysis, we only included the 477 probands who had PRD and CRD for each system.

CRD was provided by clinical geneticists or a delegate whose proforma was then checked and approved by the clinician. 69% of parents in the cohort reported the highest level of parental education at degree level or above, with 24% having achieved school leaver qualifications of some sort. Only 3% reported having not achieved any educational qualifications.

Quantity scores: PRD and CRD provide a similar quantity of data but with some notable differences by system and gene group.

PRD and CRD scores showed a similar normal distribution (figure 1a) with most providing five or six terms per participant. Not many participants had more than 16 terms provided by either source, but when this did occur, this was mostly by CRD.

PRD provided much more data for dental, gastroenterology, immunology and respiratory systems and for vision (p <0.001
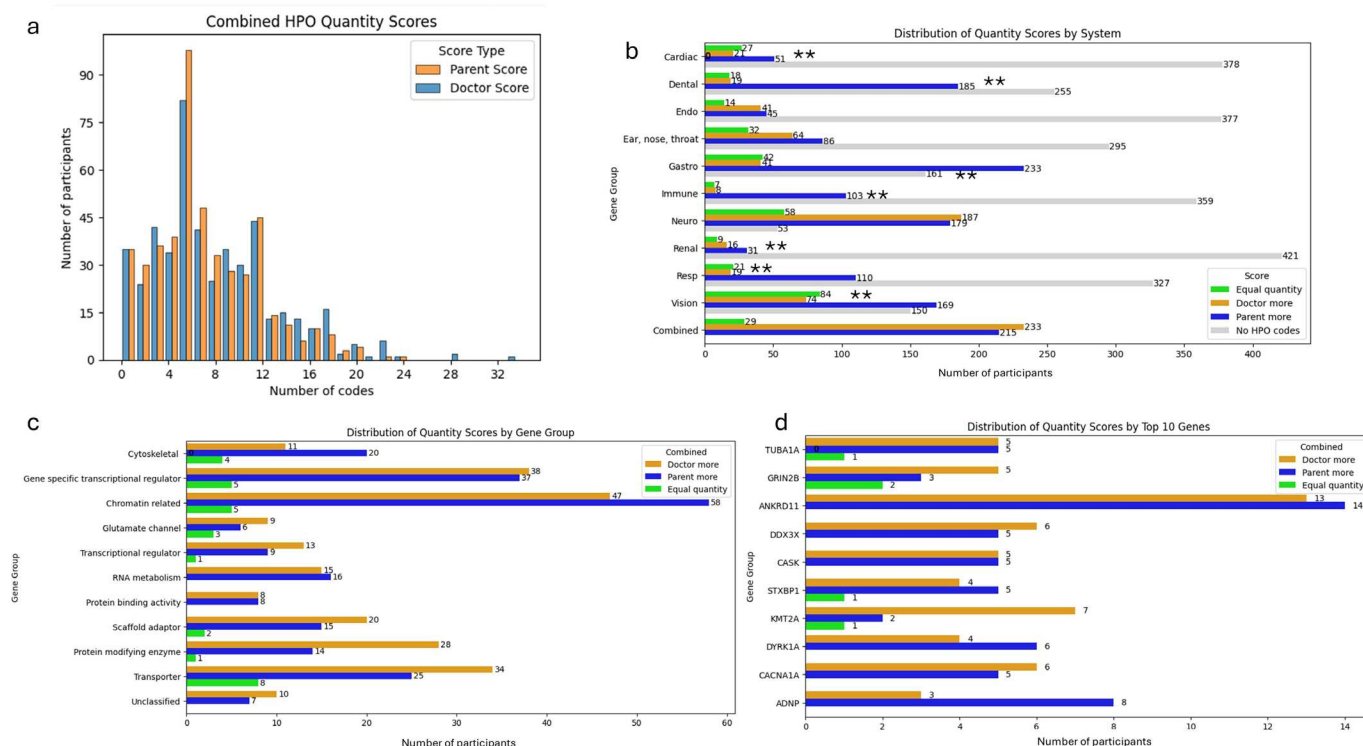
Figure 1 Quantity: (a) Histogram of combined systems HPO quantity scores. (b) Distribution of quantity scores by system. (c) Distribution of quantity scores by gene group. (d) Distribution of quantity scores for top 10 genes (highest recruited numbers in GenROC). **p value <0.05. HPO, Human Phenotype Ontology.

for all, figure 1b). This was also true for cardiac but to a lesser degree (p=0.0012). There was no evidence of a difference for endocrine, ear, nose and throat or renal. The systems with the fewest HPO codes provided (shown in figure 1b by 'No HPO codes') were cardiac, endocrine, immune and renal.

For neurology, approximately a third of the cohort had more terms provided by PRD and another third had more terms provided by CRD.

Most of the gene groups showed similar distributions for PRD and CRD (figure 1c). There was no statistical evidence of a difference in scores by gene category or by single gene (top 10 genes). Cytoskeletal and chromatin-related groups showed higher scores for PRD than CRD in comparison to transporter and protein-modifying enzyme groups which scored higher for CRD than PRD, although to a lesser extent. There were no striking differences by top 10 genes (figure 1d).

Detail scores: CRD provides more detail than PRD for most gene subgroups, combined systems and for neurology:

PRD and CRD showed similar detail by system (figure 2a). The most notable exception is for neurology, in which 60 probands had a higher detail score for CRD (p <0.001, heatmap figure 2b). This was seen to a much lesser extent in gastroenterology with 11 probands having higher scores for CRD. Vision was different to the rest of the systems with a small almost identical proportion of both higher CRD and PRD scores (28 probands had higher details scores for PRD; 24 probands had higher scores for CRD). Combined systems showed evidence for a difference with CRD having higher detail scores than PRD (p <0.001).

When looking just at the 'Parent more' and 'Doctor more' categories, all but two of the gene subgroups scored higher for detail for CRD than PRD (figure 2c), and the evidence was strongest for this the Scaffold adaptor group (p <0.01) as well as for

chromatin-related (p=0.035), gene-specific transcriptional regulator (p=0.045) and transporter (p=0.045) groups. There was no strong evidence for a difference by top 10 gene (figure 2d).

Similarity scores were low overall per participant but differed somewhat by system, gene and gene group (figure 3). Doctors provided more unique terms for neurology compared with parents who provided more unique terms for gastro, dental and respiratory systems. CRD is more similar to published syndrome phenotypes.

The scores are highest for cardiac (mean 0.75; p<0.05) and followed by gastro (mean 0.70; p<0.05), vision (0.69; p<0.05) and dental (0.63; p<0.05), while the systems that had the lowest scores were ENT (mean 0.43; p=0.8). (figure 3a). The mean similarity score for the combined systems was 0.38. Similarity scores were consistent across gene groups with a mean ranging from 0.30 (transcriptional regulators) to 0.43 in the scaffold adaptor group (figure 3b). Similarity scores were highest for STXBP1 (mean 0.5) and CACNA1A (0.49) (figure 3c). There was no evidence of a statistical difference by gene or gene subgroup.

A manual review of the 20 lowest similarity score participants showed that PRD provided terms regarding respiratory, feeding and gastroenterology, dental and autonomic features, whereas CRD described dysmorphic features and clear specific medical phenotypes in greater detail. Similarity scores were high when PRD and CRD both provided a small number of total terms and information about the same system and lowest where they provided information about different systems (see examples in figure 4).

When comparing PRD with the published HPO terms in DDG2P for each gene (figure 5), we found that half (ANKRD11, DDX3X, KMT2A, DYRK1A, ADNP) had evidence for a reduction in similarity score compared with when comparing CRD or when comparing to combined datasets. For all five of these
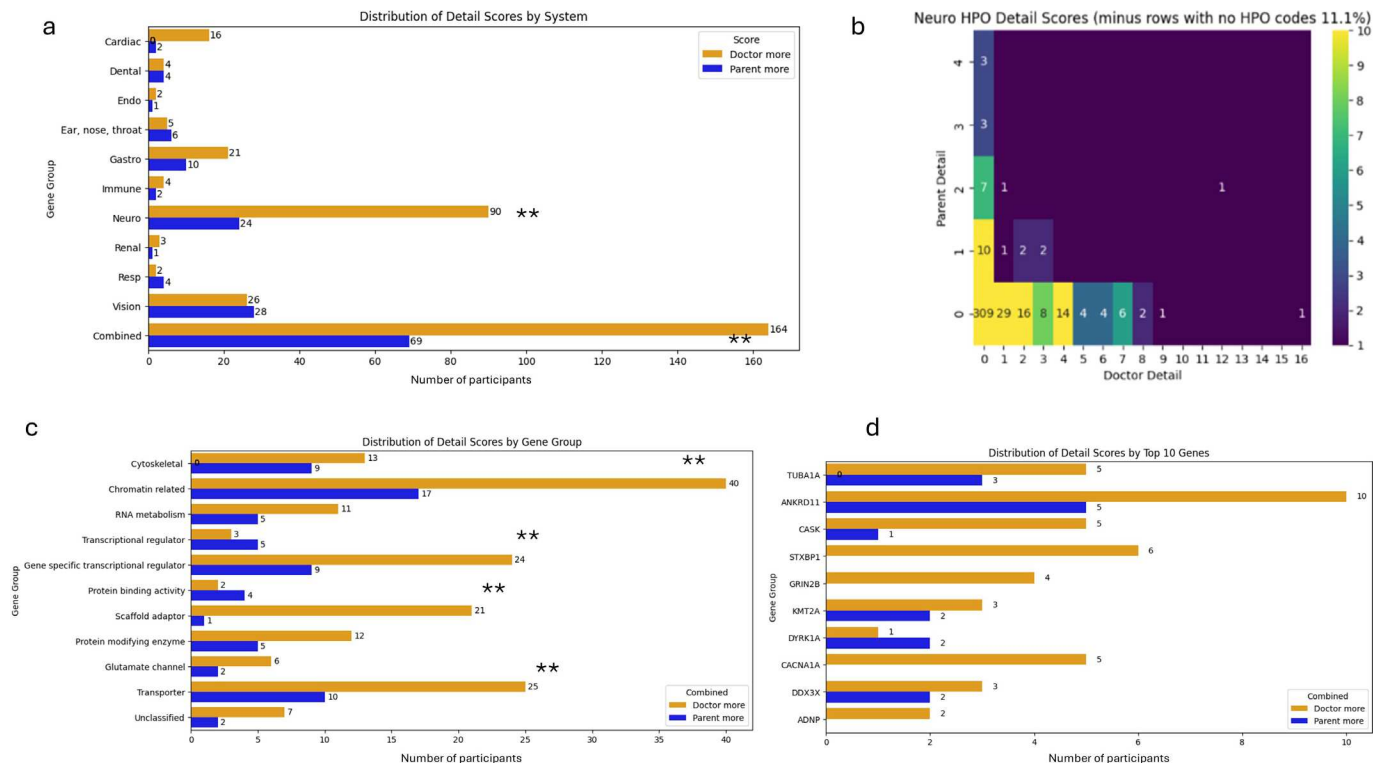
Figure 2 Detail: (a) Distribution of detail scores by system. (b) Heatmap of detail scores for neurology system. (c) Distribution of detail scores by gene group. (d) Distribution of detail scores by top 10 recruited genes. **p <0.05. HPO, Human Phenotype Ontology.
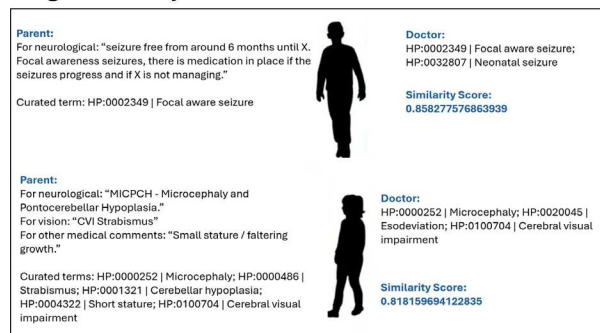
conditions, CRD is more similar to DDG2P than PRD (p <0.05). This is also true when comparing PRD and the combined datasets to DDG2P. There is no evidence for a difference between

CRD and combined for any of the genes analysed. There was no evidence of a difference between PRD and CRD with DDG2P for *TUBA1A, GRIN2B, CASK, STXBP1 or CACNA1A*. It is likely
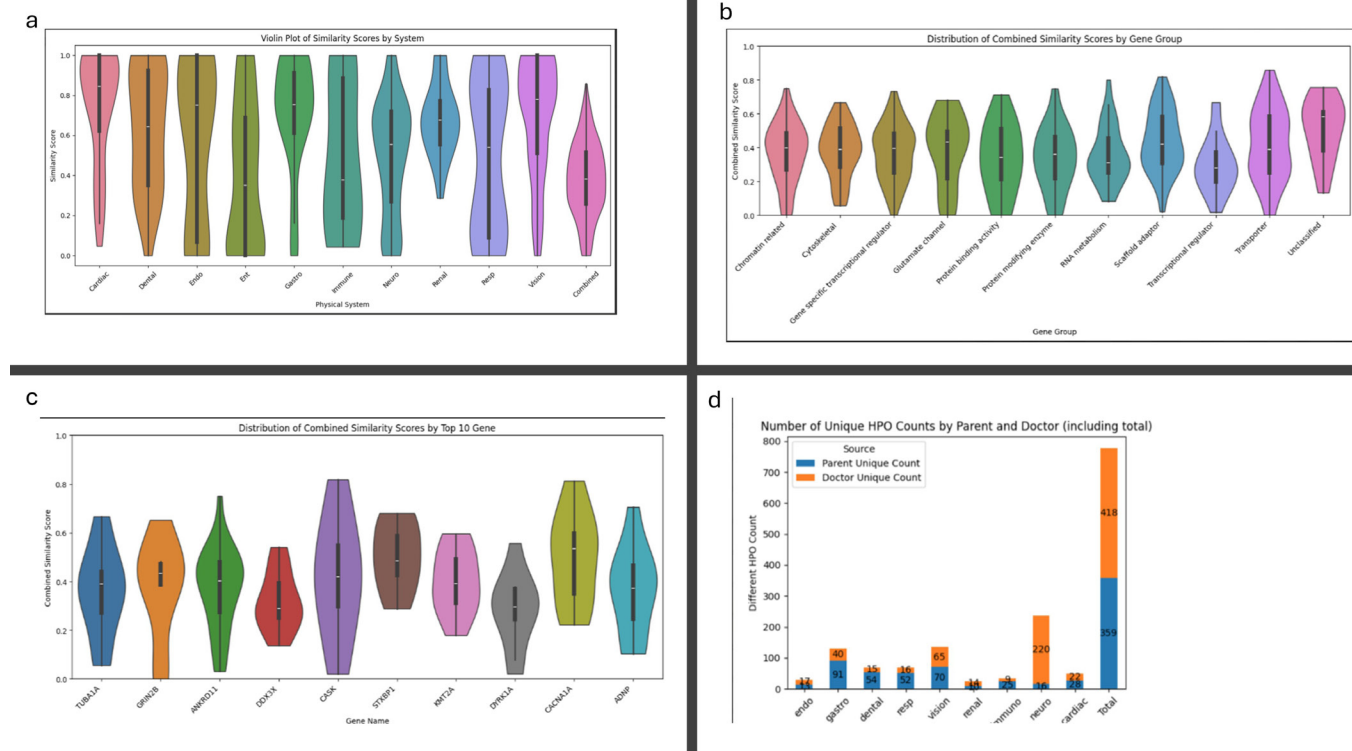
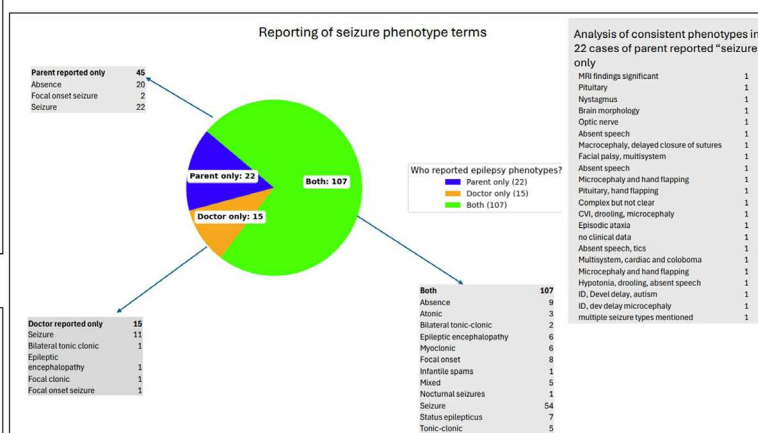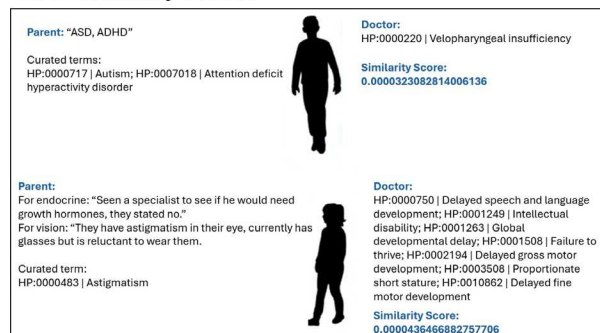**Figure 3** Similarity: (a) Violin plot with box and whisker overlay depicting similarity score distribution for all systems and depicting median and upper and lower quartiles. (b) Violin plot of similarity scores by gene groups. (c) Violin plot of similarity scores by top 10 recruited genes. (d) Number of unique HPO codes from PRD and CRD by system. CRD, clinician-reported; HPO, Human Phenotype Ontology; PRD, patient-reported.

High similarity scores

Parent:
For neurological: "seizure free from around 6 months until X. Focal awareness seizures, there is medication in place if the seizures progress and if X is not managing."

Curated term: HP:0002349 | Focal aware seizure

Doctor:
HP:0002349 | Focal aware seizure; HP:0032807 | Neonatal seizure

Similarity Score:
0.858277576863939

Parent:
For neurological: "MICPCH – Microcephaly and Pontocerebellar Hypoplasia."
For vision: "CVI Strabismus"
For other medical comments: "Small stature / faltering growth."

Curated terms: HP:0000252 | Microcephaly; HP:0000486 | Strabismus; HP:0001321 | Cerebellar hypoplasia; HP:0004322 | Short stature; HP:0100704 | Cerebral visual impairment

Doctor:
HP:0000252 | Microcephaly; HP:0020045 | Esodeviation; HP:0100704 | Cerebral visual impairment

Similarity Score:
0.818159694122835

Low similarity scores

Parent: "ASD, ADHD"

Curated terms:
HP:0000717 | Autism; HP:0007018 | Attention deficit hyperactivity disorder

Doctor:
HP:0000220 | Velopharyngeal insufficiency

Similarity Score:
0.0000323082814006136

Parent:
For endocrine: "Seen a specialist to see if he would need growth hormones, they stated no."
For vision: "They have astigmatism in their eye, currently has glasses but is reluctant to wear them."

Curated term:
HP:0000483 | Astigmatism

Doctor:
HP:0000750 | Delayed speech and language development; HP:0001249 | Intellectual disability; HP:0001263 | Global developmental delay; HP:0001508 | Failure to thrive; HP:0002194 | Delayed gross motor development; HP:0003508 | Proportionate short stature; HP:0010862 | Delayed fine motor development

Similarity Score:
0.0000436466882757706

Reporting of seizure phenotype terms

Parent reported only 45
Absence 20
Focal onset seizure 2
Seizure 22

Who reported epilepsy phenotypes?
- Parent only (22)
- Doctor only (15)
- Both (107)

Parent only: 22 | Doctor only: 15 | Both: 107

Doctor reported only 15
Seizure 11
Bilateral tonic clonic 1
Epileptic encephalopathy 1
Focal clonic 1
Focal onset seizure 1

Both 107
Absence 9
Atonic 3
Bilateral tonic-clonic 2
Epileptic encephalopathy 6
Myoclonic 6
Focal onset 8
Infantile spasms 1
Mixed 5
Nocturnal seizures 1
Seizure 54
Status epilepticus 7
Tonic-clonic 5

Analysis of consistent phenotypes in 22 cases of parent reported "seizure" only
MRI findings significant 1
Pituitary 1
Nystagmus 1
Brain morphology 1
Optic nerve 1
Absent speech 1
Macrocephaly, delayed closure of sutures 1
Facial palsy, multisystem 1
Absent speech 1
Microcephaly and hand flapping 1
Pituitary, hand flapping 1
Complex but not clear 1
CVI, drooling, microcephaly 1
Episodic ataxia 1
no clinical data 1
Absent speech, tics 1
Multisystem, cardiac and coloboma 1
Microcephaly and hand flapping 1
Hypotonia, drooling, absent speech 1
ID, Devel delay, autism 1
ID, dev delay microcephaly 1
multiple seizure types mentioned 1

Figure 4 Analysis of seizure term reporting by PRD, CRD and both. Pie chart depicting who reported epilepsy phenotypes, and for each group which subtypes were reported. For the 22 cases in which seizures were only reported by PRD, we include associated consistent neurological phenotypes for each case. CRD, clinician-reported; PRD, patient-reported.

that similarity will be higher when a small list of PRD/CRD terms is compared with a large list of DDG2P terms given the increased chance of overlap. The number of terms for each gene in DDG2P varies widely (see figure 5).

Phenotypic spread analysis showed that CRD reported more terms with more detail in neurology, but the most frequently reported terms are common to both datasets.

When assessing the number of unique phenotypes, provided there is a noticeable difference between systems with neurology alone contributing 220 of the 418 unique codes supplied by clinicians (figure 3e). Gastroenterology, dental and respiratory had more than double the number of unique codes supplied by PRD compared with CRD. Vision and cardiac had roughly equal numbers for both. Excluding developmental delay and intellectual disability, the five most frequently reported HPO terms in the combined dataset were identical to PRD top 5 and included hypotonia (198), constipation (127), seizure (119), strabismus (106) and gastro-oesophageal reflux (85). CRD overlapped but also included microcephaly (69) and hypertelorism (46).

Seizures were reported in 167 (35%) of the cohort (figure 4). Of these, seizures were reported by both sources in 64%: by CRD only in 9% and in PRD only in 27% (figure 4). For the 45 PRD-only cases, 20 had absence seizures, two had focal onset seizures and 22 just had the term 'seizure'. Manual review of the 22 'seizure' cases showed significant neurological phenotypes which would be consistent clinically with coexistent seizures in all but two cases in which data were insufficient for analysis. Parents used predominantly medical terminology specifying 'epilepsy' or 'seizures' to describe the seizures with only two using 'fits'. Two of the parents in the PRD only group used very specific terms 'focal onset seizures'.

DISCUSSION

Rare disease registries have increased exponentially in the last decade with over 800 rare disease registries listed in a 2021 European report.[28] However, while there may be an abundance, there have been concerns raised about the challenges of developing and sustaining a high-quality registry. There are large variations in the type of data being collected, which ranges from clinician-derived questionnaires[29] such as in GenIDA,[15] validated PROMS in Rare-X[16] to more basic collection tools that have not been validated, which can limit applications of the data.[30] However, rare disease data are hard to obtain and collate due to the scarcity of patients, difficulty in locating them and wide geographical spread. Clinical time is increasingly pressurised and accessing clinical data through systems remains challenging. Therefore, parent-reported registry data represent a simple, quick, inexpensive and practically achievable way to collect data and undertake research. These sorts of data seem likely to be used in the future to create datasets for machine learning and other applications. It is therefore essential to understand the strengths and weaknesses of these data. A handful of previous studies have assessed parent-reported data with the published literature and concluded that they were reasonably consistent.[8 9 11 13] However, to our knowledge, no one has directly compared parent and CRD data for consistency. Our study sought to assess this consistency on an individual level across a large cohort of NDDs to help inform the appropriate future use of such datasets.

We know from our GenROC qualitative study[31] that parents become experts in their child's condition and spend a considerable amount of time and energy on educating professionals and directing and project managing their child's care. They keep meticulous records. We also know from ThinkAloud[32] user testing of the GenROC parent proformas[17] that parents
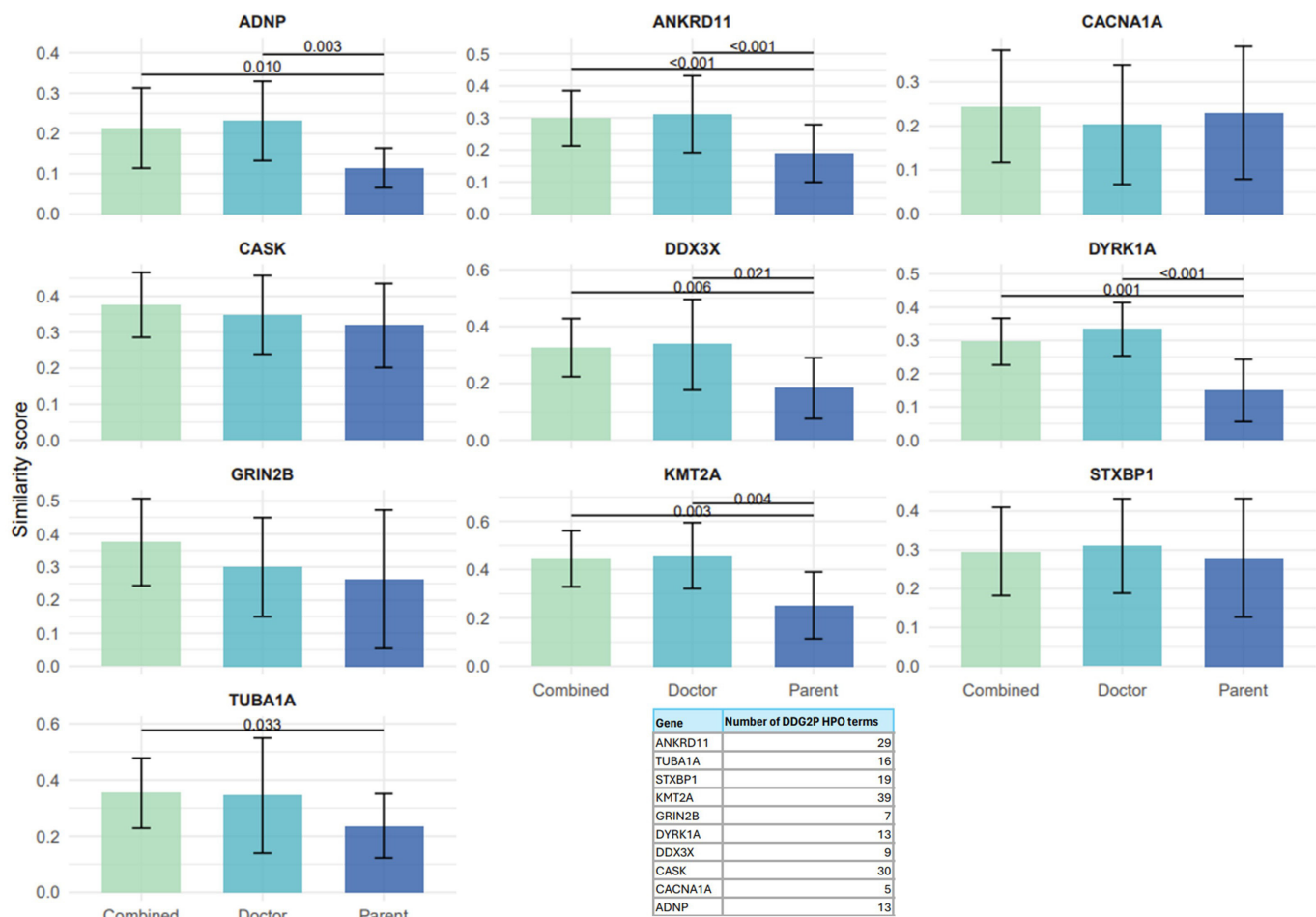
Figure 5 Similarity of reported data from parents, clinicians and combined with DDG2P published phenotypes. Clustered bar chart depicting similarity scores for each gene with DDG2P phenotypes for combined datasets, doctor only and parent only with error bars. P values are annotated where p <0.05, and the accompanying horizontal bars depict the correlating comparison (eg, combined vs parent).

will look at their child's clinic letters to help them complete the parent questionnaires. It is therefore to be expected that there should be relatively high data consistency between parents and clinicians, which is consistent with our results. We found particularly high similarity scores for the cardiac system, where most of the terms used were regarding cardiac malformations such as 'ventriculo-septal defect'. We would expect that these sorts of terms would be used in discussions with parents and in clinic letters and so would account for this level of consistency. The cardiac system also seems less likely to have as many lived experience phenotypes which might account for the similarity.

A strength of this study is that we have been able to undertake the analysis on an individual participant level comparing parent and clinician-derived data. However, a limitation of the study is that there was high missingness that resulted in exclusions. However, this could simply reflect the nature of the disorders that were included in GenROC—by definition, these were neurodevelopmental genetic disorders, and so many of these children may not have had particular problems in some of the systems (such as renal or immune), and so missing data may represent the lack of phenotype. The proportion of missing data for our analysis demonstrates the value of acquiring data from *both* parents and clinicians. By using this approach, we were able to gather data across all areas in more individuals. By requesting

data from two sources, this maximised the number of individuals for whom we had at least some data provided rather than nothing at all.

Previous studies have tried to quantify the reliability of parent-reported data by comparing parent-reported data with prevalence of phenotypes in the published literature.[11 12] These studies used a predefined generic GenIDA template in which parents were asked to respond 'yes/no/I don't know' to questions.[15] This method allowed the authors to determine phenotypes in which a difference reached 'statistical significance' such as movement disorder and hypotonia in the *DDX3X* cohort.[13] No significant differences were seen in the Koolen de Vries syndrome cohort.[12] A strength of our study compared with these previous ones is that parents were given free-text boxes which provided them with unlimited response options which could then be curated for comparison. Comparing sets of HPO terms using Most Informative Common Ancestor methods presents challenges, as the clinical relevance of the shared ancestor term can vary significantly and is not consistently correlated with its hierarchical depth within the ontology. This is likely to explain the relatively low similarity scores seen across the cohort. Nonetheless, our analysis through similarity scores using the HPO ontology allows a more detailed analysis of consistency in reporting which may be important for granularity of phenotype description especially for future applications such as for machine learning datasets.

A possible limitation of our study is that we performed a curation step in order to convert free-text parent responses into HPO terms. In order to try and minimise the introduction of clinician expertise into this step, we asked non-specialist medical students to do this work and also split the responses by system to ensure that the student could not make inferences across the dataset for an individual participant. A proportion of codes per system was double coded by a clinical geneticist to check for consistency in coding approach. While consistency was high across the curating team, a limitation exists that variability in phenotypic annotation exists even among expert clinicians, and this applies to the clinically reported data as well.

Another limitation of this study is the skewed nature of parental educational experience that is not representative of the general population.

Clinical proformas were completed by clinical geneticists (or a delegate). These are busy clinicians who may have had very limited time to provide this data. We therefore noted differences between sites and clinicians with respect to the 'fullness' of the clinician proforma. In addition, some sites used an appropriately trained research nurse, genetics trainee or genetic counsellor to complete the proforma. In those situations, they would also spend time consulting wider hospital records to ensure they were able to provide as comprehensive a set of answers as possible. This is unlikely to have been possible for the senior clinicians who had significant time restrictions due to competing clinical priorities. As such, it is likely that there could be a considerable difference in scores if it had been possible to perform sub-analyses based on data collectors. Nonetheless, this reflects real-life data collection in clinical sites—there is likely to be variability in data reporting in any phenotyping study dependent on who is providing the data both in terms of experience, seniority and skill but also just as importantly in terms of allocated time for the task and motivation to do so.

Clinical geneticists take a holistic approach which would include a general developmental and paediatric history. Nonetheless, not all areas may be covered, and indeed parents may not mention everything as they may think it not relevant to the consultation, especially where time pressure is an issue. Clinicians may not have seen the child for a period of time in which a new feature may have arisen. This may result in clinicians simply not knowing or not being aware of issues in certain areas. This could account for the evidence for a difference in quantity score seen in the immune, dental, cardiac, gastroenterology, renal, vision and respiratory systems. Parents are likely to be able to report more in these areas than the clinical geneticist who may have only seen the child once or twice and may be unaware of some of these issues. This has also been reported elsewhere in a study of caregiver-reported information in 237 individuals with Koolen de Vries syndrome which identified that childhood asthma and recurrent pneumonia are respiratory features seen in 40% of this group and not reported previously in clinically reported cohorts.[12]

Unique codes from parents in GenROC included nasal congestion, recurrent respiratory infection, asthma and loud snoring for the respiratory system and gingival bleeding, fragile teeth and delayed and advanced eruption of teeth and tooth abscess for the dental systems. This represents a strength of parent-reported data in that it represents lived experience data which is likely to be different in granularity to clinician data.[13 15]

Some of these reported terms may reflect common paediatric conditions that are likely unrelated to the genetic diagnosis (see online supplemental table 2), but some may prove to be an important lived experience phenotype. In our cohort, constipation

is the second most frequently provided parent term (26%), which is much higher than that reported in the general paediatric population (about 3% worldwide).[33] Strabismus is reported in 14% in GenROC compared with 2%–2.5% in the paediatric population.[34] In contrast, myopia was reported in 6% in GenROC compared with 16% in the wider paediatric population,[35] with the caveat that it is difficult to accurately compare like to like. It is therefore important to collect lived experience phenotypes, but this must be compared with general paediatric prevalence before any definite syndromic link can be inferred. This can be difficult as data are sparse and vary widely regarding these common presentations.

The neurological system was notably different from the rest with evidence of a difference in detail scores along with the large number of unique codes in CRD. It is likely that this is due to the cohort being made up of children with NDDs and therefore a high proportion of neurological phenotypes. Clinicians are likely to therefore have and provide the most information about this system given it may be seen as most relevant to the genetic diagnosis. When reviewing the list of unique codes for neurology for the clinicians, these are made up of MRI Brain descriptors (such as polymicrogyria or cerebellar atrophy), seizure phenotypes (such as focal motor seizure and absence seizure) and clinical phenotypes (such as ataxia, tremor and dystonia). It is plausible that some of these terms at least may not be known to parents or that parents may provide some but not all of them when responding. It is also possible that some parents may assume that CRD will provide this data for their child and so may not have provided it all. It is interesting, however, that the genes with the highest similarity scores were *CACNA1A* and *STXBP*1. These genes represent fairly tightly defined neurological phenotypes and may explain this higher consistency compared with the rest.

Comparison of CRD and PRD with published phenotypes in DDG2P for *ANKRD11, DDX3X, KMT2A, DYRK1A* and *ADNP* showed that CRD was more similar to DDG2P than PRD. PRD did not increase the similarity if added to CRD(4d). This is perhaps not surprising, though, given the DDG2P list is curated from the published literature which currently largely consists of clinically reported phenotypes. This difference was clear for these genes that are more syndromic and multisystem in nature as compared with the more tightly defined predominantly neurological phenotype genes *TUBA1A, GRIN2B, CASK, STXBP1* and *CACNA1A* in which we found no evidence for a difference between CRD, PRD or combined when generating similarity scores with DDG2P.

CONCLUSIONS

Parents report a similar amount of data to clinicians, but the content differs and is more likely to represent common childhood phenotypes. Clinician-reported data is likely to miss that lived experience granularity but is more likely to contain specific clinical phenotypes, for example, MRI findings which may be missed or may not be known to parents. The gold standard would be to include both data sources, but this is unlikely to be the case for most studies due to time, cost and resource implications. These nuances should be considered by researchers who are setting up studies or who may be choosing to use specific datasets and when considering the strengths and limitations of said datasets for future applications. PRD could be improved through use of standardised co-produced user-tested measures specifically validated for rare NDDs. Automated tools designed to accurately convert free text to HPO using machine learning assistance could be helpful in the future. Understanding the use of the intended data, such as for genotype-phenotype inference or natural history progression, is critical to the design and choice of the questionnaires and measures to ensure they are fit for purpose for both research and for measuring what is important to patients

and their families. Lived experience data should be included in future publications to ensure fullness of phenotype descriptions.

ORCID iDs
Karen Jaqueline Low https://orcid.org/0000-0002-4975-9363
Mevmi Lasanya Kodippuli Thanthilla https://orcid.org/0009-0006-3322-8329

REFERENCES

1 Witt S, Schuett K, Wiegand-Grefe S, *et al*. Living with a rare disease - experiences and needs in pediatric patients and their parents. *Orphanet J Rare Dis* 2023;18:242.
2 Wright CF, Campbell P, Eberhardt RY, *et al*. Genomic Diagnosis of Rare Pediatric Disease in the United Kingdom and Ireland. *N Engl J Med* 2023;388:1559–71.
3 Chung WK, Herrera FF. Health supervision for children and adolescents with 16p11.2 deletion syndrome. *Cold Spring Harb Mol Case Stud* 2023;9:a006316.
4 The Simons VIP Consortium. Simons Variation in Individuals Project (Simons VIP): A Genetics-First Approach to Studying Autism Spectrum and Related Neurodevelopmental Disorders. *Neuron* 2012;73:1063–7.
5 Kingdom R, Wright CF. Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts. *Front Genet* 2022;13:920390.
6 Vasireddi SK, Draksler TZ, Bouman A, *et al*. Arrhythmias including atrial fibrillation and congenital heart disease in Kleefstra syndrome: a possible epigenetic link. *Europace* 2023;26.
7 Low KJ, Walker M, Treneman-Evans G, *et al*. Life Beyond Childhood: Insight Into the Lived Experience of 91 Adults With KBG Syndrome Through an Online Patient/Caregiver-Reported Co-Produced Questionnaire. *Brain Behav* 2025;15:e70553.
8 Engwerda A, Frentz B, den Ouden AL, *et al*. The phenotypic spectrum of proximal 6q deletions based on a large cohort derived from social media and literature reports. *Eur J Hum Genet* 2018;26:1478–89.
9 Engwerda A, Frentz B, Rraku E, *et al*. Parent-reported phenotype data on chromosome 6 aberrations collected via an online questionnaire: data consistency and data availability. *Orphanet J Rare Dis* 2023;18:60.
10 Engwerda A, Kerstjens-Frederikse WS, Corsten-Janssen N, *et al*. The phenotypic spectrum of terminal 6q deletions based on a large cohort derived from social media and literature: a prominent role for DLL1. *Orphanet J Rare Dis* 2023;18:59.
11 Bain JM, Snyder LG, Helbig KL, *et al*. Consistency of parent-report SLC6A1 data in Simons Searchlight with Provider-Based Publications. *J Neurodev Disord* 2022;14:40.
12 Colin F, Burger P, Mazzucotelli T, *et al*. GenIDA, a participatory patient registry for genetic forms of intellectual disability provides detailed caregiver-reported information on 237 individuals with Koolen-de Vries syndrome. *Genet Med Open* 2023;1:100817.
13 Ruault V, Burger P, Gradels-Hauguel J, *et al*. Lessons from two series by physicians and caregivers' self-reported data in DDX3X-related disorders. *Mol Genet Genomic Med* 2024;12:e2363.
14 Zdolšek Draksler T, Bouman A, Guček A, *et al*. Exploring Kleefstra syndrome cohort phenotype characteristics: Prevalence insights from caregiver-reported outcomes. *Eur J Med Genet* 2024;72:104974.
15 Burger P, Colin F, Strehle A, *et al*. GenIDA: an international participatory database to gain knowledge on health issues related to genetic forms of neurodevelopmental disorders. *J Neural Transm (Vienna)* 2023;130:459–71.
16 Trzupek K, Bhargava R, Kuan C, *et al*. Breaking barriers in rare disease research: The RARE-X Open Science Data Challenge as a model for collaborative innovation and community partnership. *HGG Adv* 2025;6:100462.
17 Low KJ, Watford A, Blair P, *et al*. Improving the care of children with GENetic Rare disease: Observational Cohort study (GenROC)-a study protocol. *BMJ Open* 2024;14:e085237.
18 Harris PA, Taylor R, Minor BL, *et al*. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform* 2019;95:103208.
19 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
20 Foreman J, Brent S, Perrett D, *et al*. DECIPHER: Supporting the interpretation and sharing of rare disease phenotype-linked variant data to advance diagnosis and research. *Hum Mutat* 2022;43:682–97.
21 Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 2013;41:D377–86.
22 Thomas PD, Ebert D, Muruganujan A, *et al*. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci* 2022;31:8–22.
23 Lachenbruch PA. McNemar test. wiley statsref: statistics reference online.
24 Abramovich F, Benjamini Y. False discovery rate. Wiley StatsRef Statistics Reference Online.
25 Ridsdale AM, Dickerson A, Chundru VK, *et al*. Phenotypic spectrum of dual diagnoses in developmental disorders. *Am J Hum Genet* 2024;111:2382–91.
26 Welch BL. On the comparison of several mean values: An alternative approach. *Biometrika* 1951;38:330–6.
27 Ruxton GD. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behav Ecol* 2006;17:688–90.
28 Orphanet. Rare diseases registries in Europe. 2021.
29 Bayat A, Grimes H, de Boer E, *et al*. Natural history of adults with KBG syndrome: A physician-reported experience. *Genet Med* 2024;26:101170.

30 Hageman IC, van Rooij IALM, de Blaauw I, *et al*. A systematic overview of rare disease patient registries: challenges in design, quality management, and maintenance. *Orphanet J Rare Dis* 2023;18:106.

31 Low KJ, Treneman-Evans G, Wynn SL, *et al*. Identifying the Impacts, Obstacles and Information Barriers for Parents of Children Living With Genetic Neurodevelopmental Disorders: A Qualitative Study. *Health Expect* 2025;28:e70340.

32 Bradbury K, Morton K, Band R, *et al*. Using the Person-Based Approach to optimise a digital intervention for the management of hypertension. *PLoS One* 2018;13:e0196868.

33 Tabbers MM, DiLorenzo C, Berger MY, *et al*. Evaluation and Treatment of Functional Constipation in Infants and Children. *J pediatr gastroenterol nutr* 2014;58:258–74.

34 Williams C, Northstone K, Howard M, *et al*. Prevalence and risk factors for common vision problems in children: data from the ALSPAC study. *British Journal of Ophthalmology* 2008;92:959–64.

35 Rudnicka AR, Kapetanakis VV, Wathern AK, *et al*. Global variations and time trends in the prevalence of childhood myopia, a systematic review and quantitative meta-analysis: implications for aetiology and early prevention. *Br J Ophthalmol* 2025;43.