1-1-2023

"Come On. I Need An Answer." A Mixed-Methods Study Of Barriers And Disparities In Diagnostic Odysseys

Zeyu Tang

“Come on. I need an answer.”

A mixed-methods study of barriers and disparities in diagnostic odysseys

A Thesis Submitted to

the Yale University School of Medicine

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Medicine

by

Zeyu Tang

2023

Abstract

A mixed-methods study of barriers and disparities in diagnostic odysseys

Background: The boom of next generation DNA sequencing over the past decade has improved our ability to provide accurate genetic diagnoses for children with previously undiagnosed diseases, in turn leading to important advances in management and prognostication. Even given this progress, two areas of ongoing need are the accurate definition of further novel genetic diseases and to make genetic expertise and diagnostics widely available to children and families who have frequently endured grueling diagnostic odysseys. The Pediatric Genomics Discovery Program (PGDP) at Yale is an advanced genomics program focusing on both these areas, enrolling over 700 patients since its inception and eventually providing approximately one-third with new genetic diagnoses. Despite this success, we questioned whether the PGDP was achieving its full potential for impact by reaching a broad, representative participant population.

Hypothesis: Current PGDP participant demographics are not representative of the racial/ethnic and socioeconomic diversity in the community of patients with potentially undiagnosed genetic diseases, which may relate to systemic barriers along the diagnostic odyssey.

Methods: We created a questionnaire and in-depth interview process for existing PGDP participants to evaluate barriers to diagnostic care, then analyzed transcripts for themes. We analyzed demographic characteristics and referral routes of the PGDP cohort to find factors related to recruitment. We developed a screening tool based on diagnostic codes and queried the Yale New Haven Health System (YNHHS) electronic health record (EHR) to identify inpatient children between 2017-2022 with potentially undiagnosed genetic conditions, estimate their prevalence, and compare their characteristics with those already enrolled in PGDP. Then, we manually reviewed patient charts further narrow patients down to those who likely had undiagnosed genetic diseases. We used Pearson chi-square for categorical data, a multinomial regression model for predictors of enrollment, and Kruskal-Wallis one-way analysis of variance with pairwise comparisons with Bonferroni correction for multiple comparisons.

Results: Survey results noted 1) Not knowing the PGDP existed (42%) and 2) Not knowing if they qualified for PGDP (36%) as the most common barriers to participant enrollment. Qualitative interviews identified three overarching themes related to the search for a unifying

medical diagnosis for patients and families: 1) Challenges along the diagnostic odyssey (largely barriers in the healthcare system), 2) Tools to navigate the uncertainty (particularly parent serving as a care-captain) and 3) Perceptions of the PGDP (having reservations about participating vs desire for a diagnosis). In the PGDP cohort analysis, being directly identified by a PGDP-affiliated physician was associated with the highest representation of URM (52%) compared to referrals through Yale Genetics (27%) or Other Referrals (16%), and a significantly greater URM representation compared to both the national pediatric population ($p=0.008$) and to a peer genetics program ($p<0.0001$). Our EHR screen identified 1,648 potential undiagnosed inpatients, of which 36 were known PGDP participants. Manual chart review of the first 162 confirmed that 71 (44%) were appropriate candidates for evaluation of a potentially undiagnosed genetic disorder. These potentially undiagnosed patients had a significantly higher proportion of Black (24%) and Hispanic (31%) children relative to New Haven county ($p=0.033$).

Conclusions: Our findings showed that significant barriers exist to reaching the PGDP, from challenges with the diagnostic odyssey to simply being aware that such programs exist. Families emphasized benefits of parents acting as care-captains, implying the need for time and other resources as well as for an understanding of how to navigate the health system. Not unexpectedly, referral routes requiring greater distances (physical or through the health system) to reach the PGDP were dominated by White participants, likely reflecting the difficulty of URM to overcome barriers. Our EHR screening tool allowed us to find a large number of previously unknown undiagnosed patients that would be appropriate candidates for the PGDP. Tellingly, this group had over-representation of URMs, providing further evidence that these patients are being missed by traditional routes of referral. Overall, these data deliver valuable insights that are actively being used to alter the PGDP's approach to participant enrollment and are also applicable more broadly to other programs struggling with similar disparity challenges.

Acknowledgements

First and foremost: Dr. Saquib Lakhani. Thank you for your generous mentorship over this past year. All your encouragement, thoughtfulness, and support made this research year the foundation for life-long learning. You are what is so extraordinary about Yale. I am - and forever will be - a diehard fan.

Dr. Lauren Jeffries, Monica Konstantino, and the rest of the PGDP crew: Your feedback and assistance in this work has been invaluable. I hope my work will help families on the diagnostic odyssey in the years to come. Drs. Nicola Hawley and Paul Aronson, thank you for your time and considerate expertise. Your methodological guidance gave me the tools and confidence to try new things.

Thank you, Office of Student Research. Your support has made this research possible. Kelly Jo Carlson, thank you for being there during the (several) times I needed your help these last five years.

Family, friends, and Elizabeth Lee: Thank you, thank you, thank you. You mean the world to me.

TABLE OF CONTENTS

INTRODUCTION

The burden of rare diseases

Although there is no broadly accepted definition of a rare disease, the Rare Disease Act of 2002 describes diseases that affect "populations smaller than 200,000 individuals in the United States,"[1] or roughly 1 in 1,500, with the European Union using a similar prevalence of 1 in 2,000 or fewer.[2] This means that any single rare disease may only be seen once or twice in a lifetime for most clinical practitioners; however, to think of rare diseases as isolated problems misses the forest for the trees. If we add the prevalence of individual conditions together, we see that rare diseases as a group are estimated to affect 3.5-5.9% of the general population or over 400 million people worldwide.[3,4] Not only are individuals with rare diseases common as a collective, they have a disproportionate impact on healthcare utilization. One 2016 study found that children with rare diseases represented 20% of all pediatric hospitalizations in the U.S., and cost $105 billion – 50% greater than all other children combined; they had sixfold higher hospital costs, threefold longer length of stay, and thirteenfold higher mortality rate compared to other children.[5] Patients with rare diseases have similarly difficult, expensive, and extensive experiences navigating the healthcare system.[5-7]

Next-generation DNA sequencing: a breakthrough for diagnostics

Orphanet,[8] the largest database of rare diseases, contains a catalogue of over 10,000 rare diseases along with over 8,000 disease-gene relationship, suggesting that roughly 80% of rare diseases have an underlying genetic etiology and emphasizing the

importance of genetic testing for accurate diagnosis and management. Next generation DNA sequencing (NGS) approaches, which take advantage of massively parallel sequencing techniques combined with efficient bioinformatic tools, have revolutionized genomic medicine by dropping the cost of sequencing the entire genome from the original $2.7 billion in to the current cost of around $500. After the first successful clinical diagnosis by exome sequencing (ES) was made in 2009,[9] NGS techniques such as ES have been increasingly incorporated into the standard genetic diagnostic work-up. Clinical ES compares a patient's exome – the protein coding regions of the entire genome – to a database of ~8,000 disease-associated genes and their known pathogenic variants to find a suitable diagnosis.[10] Currently, these methods are used when karyotypes, microarrays, or targeted gene panels fail to diagnose suspected genetic conditions, though ES and even genome sequencing (GS) – which sequences the entire genome, both coding and non-coding regions – are increasingly being used as a first-line early diagnosis tools, especially for critically ill children when rapid diagnosis can quickly guide crucial management decisions.[11]

Genomic medicine is still in its infancy, however, and even our most comprehensive clinical tests pale in comparison to what we do not know. Only in 2022 did we manage to sequence the remaining 8% of the human genome[12] and of the approximately 21,000 currently defined genes, 20% still have no clear function.[13] Although clinical ES only looks at the approximately 8,000 genes with known disease-gene correlations, the rapid pace of ongoing discovery of gene-phenotype correlations (**Figure 1**)[14] suggests that the number of known disease-causing genes will continue to increase in the coming years. As an illustration, the most recent 2022 update of the

International Union of Immunological Societies Expert Committee reported a total of 485 genetic causes of immune dysfunction, with *55 novel genetic causes* defined or confirmed/expanded in the past two years alone.[14]

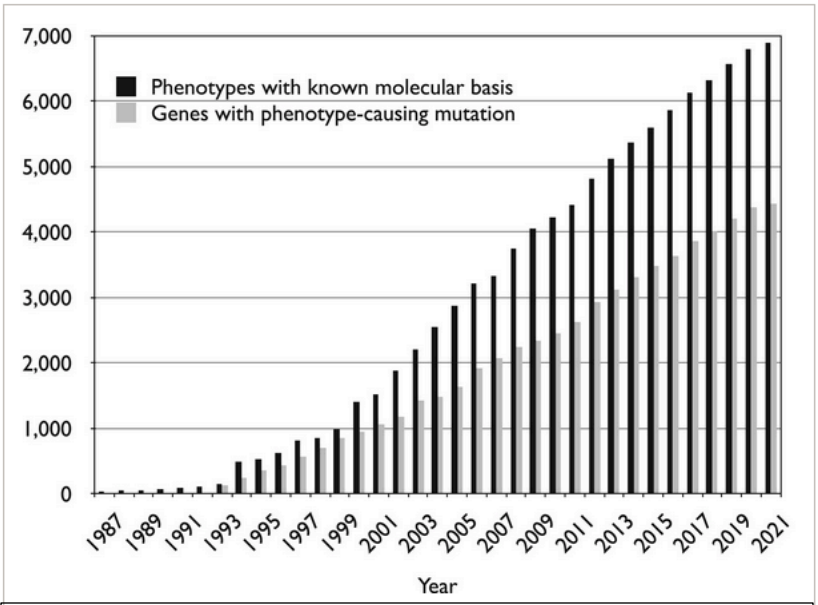

Figure 1: Pace of new gene-phenotype correlations. Graph (taken from Hamosh, et al., 2021) showing pace of new gene-phenotype correlations found in Online Inheritance in Man (OMIM), one of the key resources for such information.

Thus, NGS serves as an invaluable tool for both diagnosis of known disease-causing genes in the clinical setting and for discovery of novel gene-phenotype relationships in the research setting. Still, despite the remarkable advances made possible by NGS, many patients and families endure numerous challenges in reaching a clear diagnosis, whether due to delays in recognizing a known rare genetic disease or hurdles related to a previously undescribed condition.[15,16]

The diagnostic odyssey

Although a child with congenital heart disease may be different from one with epileptic encephalopathy in presentation, mechanism, and genetic defect, they may nonetheless share similar medical experiences in search of diagnosis – the diagnostic odyssey. The diagnostic odyssey is a prolonged, wandering search for a diagnosis across medical specialists and programs.[17] Wanderers embark on this quest for many reasons. They hope that at the end of their odyssey, they will find an explanation for

their health problems; validation, closure and social connection; informed changes to their current care; predictions about their future; access to new services or treatments.[17,18] Over time, a portion of the travellers may find a diagnosis through genetic testing, while others may stop due to resources, motivation, or death. The 2020 Annual Impact Report by Global Genes[19] found that overall, patients took an average of seven years to obtain an accurate diagnosis, and required an average of seven specialists.[19]

But active seekers are only a small cohort of the undiagnosed within diagnostic odyssey – the pediatric population as a whole is under-diagnosed. Children with diagnosed or suspected genetic disease in the U.S. account for 2.6-14% of all pediatric hospital admissions, but are responsible for 11-46% of the total cost of all pediatric hospitalizations in the nation; critically ill newborns have by far the highest healthcare cost per capita out of all pediatric groups.[20] Half of all critically ill newborns in the U.S. have a suspected genetic disease.[20] Thus, any improvements to this vulnerable population will likely lead to healthcare improvement. In fact, ES has been shown to be most cost-effective first-line diagnostic tool for critically ill newborns in terms of diagnostic yield,[21] cost per diagnosis[21,22] and cost per quality-adjusted-life-years.[21] But in reality, only 7.6% of critically ill newborns ever receive any genetic testing during admission, and the average time taken to first test is well over four months.[23] Moreover, racial and ethnic disparities exist in newborn access to basic genetic evaluations.[24] Children with potentially undiagnosed genetic conditions are a high-cost cohort to the healthcare system, and there is a major unmet need for genetic

diagnosis. Early genetic diagnosis - in other words, shortening the diagnostic odyssey[25] – is a key component in health systems quality improvement efforts.

The need for ongoing investigation into rare genetic diseases has led to the development of multiple advanced genomic diagnostic programs including the Undiagnosed Diseases Network (UDN) through the National Institutes of Health in the United States, the 100,000 Genomes Project through Genomics England and the National Health Service in the UK, the Undiagnosed Diseases Program in Western Australia.[17,26,27] Collaborative open-access data-sharing networks such as GeneMatcher have further accelerated the rate of diagnosis and discovery by matching undiagnosed patients with similar phenotypes or gene defects to other patients around the world.[28]

Yale University's Pediatric Genomics Discovery Program (PGDP) was the primary focus of this thesis. PGDP evaluates undiagnosed children with a broad range of phenotypes using NGS to first find diagnostic or candidate genes/variants, followed by functional basic science experiments to validate these candidates with the ultimate goal of providing a molecular genetic diagnosis.[27] Importantly, the PGDP engages participants in the program by returning both genetic and basic science research results, which is incorporated into the IRB protocol, thereby giving patients and families an active role in their own journeys. This commitment to individual participants, unlike research programs with deidentified human subjects, creates a close partnership between patients, families, and the PGDP team.

In the 7 years since its inception, PGDP has evaluated over 700 patients through direct enrollment and collaborations with other investigators (internal data). The most

common phenotypes are Immune System defects and Syndromic/Multisystem disease (each account for 29% of the total cohort) followed by Cardiovascular (17%) and Nervous System (11%) abnormalities.[27] The overall rate of successful diagnosis is approximately 21%,[27] which is in line with other advanced genomics discovery programs.[29] Additionally, insights from these patients have allowed for the discovery of 22 new variants of existing genes and one new gene without a previously known association with disease.[27]

The PGDP's definition of an undiagnosed patient is someone who has signs or symptoms that greatly raise clinical suspicion for an underlying genetic disorder, without a definitive molecular diagnosis.[10,27] This definition is purposely broad in order to avoid false negatives, and encompasses patients with a variety of clinical presentations without a known underlying explanation including fetal or neonatal death, congenital malformations or syndromic appearance, developmental delay, or immune dysfunction. Patients with a family history similar unexplained signs or symptoms are particularly good candidates for evaluation.

The PGDP generally has three sources of patients with varying levels of interaction with the health system. One group of patients are those actively followed by a geneticist and have received negative clinical testing. The second group comes due to referrals from specialists, or even self-referrals. The third group are PGDP-affiliated pediatric critical care physicians who see potentially undiagnosed patients on the wards. In general, interested patients and families do not require any specific, formal referrals or applications to be evaluated. Notably, the program welcomes those that might otherwise be "missed" by the healthcare system. This includes those who have

never seen a genetics specialist and those who have been genetically evaluated but prior to the age of ES with its significant improvements in diagnostic yield.

It is worth mentioning here that traditionally URM, namely Hispanic/Latinx and African American communities are more likely to remain undiagnosed compared their White non-Hispanic counterparts,[30,31] and this is due to disparities across several areas. URM lack a basis for “normal” genetic results, lack access to clinical genetics services, and remain under-represented in genetics research.[32] With a lower diagnosis yield, we would expect a disproportionate number of families in the diagnostic odyssey who are URM. But instead, we see the opposite: applications[18] and accepted cohorts[29] for U.S. advanced diagnostic programs demonstrate lower proportions of URM as compared to what would be expected based on the demographics of the U.S.

STUDY PURPOSE

Next generation sequence has revolutionized medical care for many patients and families by providing clear genetic molecular diagnoses for their conditions, leading to improvements in diagnosis, management, and counseling. However, the benefits of the NGS revolution have not been felt equally across various groups with well-documented disparities for underrepresented minorities (URM). The Yale Pediatric Genomics Discovery Program (PGDP) is a rare disease program with an overarching goal of helping undiagnosed patients and families bring an end to their diagnostic odysseys. Despite its success in this mission, the PGDP has not previously assessed the challenges faced by families in reaching the PGDP and whether certain groups are disadvantaged in accessing the program.

We hypothesized that the current participant demographics in the PGDP are not representative of the diversity in the community of patients with undiagnosed genetic diseases, at least in part due to systemic barriers along the diagnostic odyssey.

By addressing this hypothesis, we aim to improve access to advanced genetic diagnostic services for patients and families navigating through the diagnostic odyssey. Identifying disparities and associated barriers to access will allow us to reach more families that might otherwise never receive a genetic diagnosis.

SPECIFIC AIMS

Aim 1: Understand how patients and families navigate barriers in the diagnostic odyssey

We used surveys and in-depth semi-structured interviews to gain a broad understanding of barriers in the diagnostic odyssey, and how families navigated them to reach the PGDP.

Aim 2: Investigate demographics of PGDP participants based on route of recruitment

Through analysis of PGDP cohort data, we investigated the three primary routes of access to PGDP and how each route related to racial/ethnic and socioeconomic demographics of the participants.

Aim 3: Identify children with potentially undiagnosed diseases in the Yale-New Haven Health System

To find and characterize children with potentially undiagnosed genetic diseases, we designed a screening tool based on diagnostic codes and applied it to inpatient admissions in the YNHHS electronic health record.

STUDY AIM #1: UNDERSTAND HOW PATIENTS AND FAMILIES NAVIGATE THE BARRIERS OF THE DIAGNOSTIC ODYSSEY

Introduction

The diagnostic odyssey is a common experience for patients and families with undiagnosed conditions in which they navigate the healthcare system in search of a diagnosis for their condition.[17] We wanted to understand what families experience from the first moment they realize something is wrong to the moment they reach the PGDP. While the feelings of stigma, lack of knowledge among healthcare professionals, and lack of social support have been described in general terms for rare diseases[33] (and, to a limited extent, undiagnosed patients in the U.S.),[18,34,35] scant information is available focusing on the barriers to access for tangible resources that families face along the journey and how successful families navigate these challenges. Similar genomic diagnostic programs in the same institution can have very different racial/ethnic, education, financial, and cultural characteristics depending on hospital department, which indicates that some barriers selectively affect certain groups more than others.[15] To truly understand the experiences of undiagnosed families, we wanted to focus on barriers, understanding what specifically hold families back, and what tools families use to navigate these challenges to move forward in the diagnostic odyssey.

Methods

Study population and recruitment

The categories of patients within the PGDP cohort have previously been described.[27] Briefly, our study population was the 208 identifiable participants recorded in the PGDP under research protocol HIC: 1411014877 from 2015 to 2022 who had individually identifiable demographic information. Enrolled participants are those who have signed a written informed consent form to participate in the program and have provided genetic samples for analyses; Not Enrolled participants completed the consent form for participation in PGDP but did not progress further into the program because they declined to provide a genetic sample, were deemed ineligible based on missing parameters necessary for genetic analyses, or were lost to follow-up.

Of participants with identifiable demographic information, 175 had deliverable e-mail addresses and were included in our study. The survey and interviews were conducted between August 30, 2022 to December 1, 2022. Bilingual survey invitations (English and Spanish) were e-mailed to patients via their listed addresses in the PGDP database. Follow-up reminder e-mails were sent out on a bi-weekly basis to those who had not completed the survey or left them incomplete. We distributed $10 Amazon gift card vouchers to each participant upon survey completion.

Survey

Our survey asked participants for their demographics and significant barriers along the diagnostic odyssey. We defined a barrier as any factor that delayed the

respondent between the beginning of their journeys until initial evaluation by PGDP. We asked participants to evaluate the significance 16 barriers along three different areas: access to healthcare services or specialists, mistrust or concerns about the program, and logistical problems (**Appendix A**). Participants were asked to estimate the significance of each barrier along a Likert scale (**Table 2**). A minor barrier was defined as a barrier that delayed PGDP enrollment by less than 7 days, a moderate barrier was a delay of weeks to months, and a major barrier was one that delayed enrollment by months to years. After, respondents were asked to rank the barriers they described as moderate or major in order of significance. We hosted our bilingual survey on the Qualtrics XM™ (Provo, UT, USA) platform.

1	2	3	4
Not a barrier or This did not cause any delays for me	A minor barrier or This may have delayed me by less than 7 days	A moderate barrier or This may have delayed me by less than 4 weeks	A major barrier or This may have delayed me by months to years

Table 2. Likert scale for respondents to estimate significance of barriers.

Interviews

At the end of the survey, participants could opt-in to a follow-up interview for an additional $50 Amazon gift card voucher. Since our main concern was access for African American and Hispanic patients, we created quotas of Enrolled vs Not Enrolled in PGDP, with groups evenly divided in groups of ten according to racial/ethnic groups of Hispanic/Latinx, African American non-Hispanic/Latinx and All other non-Hispanic/Latinx patients. Our goal was thus 60 total participants. The interviews were

semi-structured and focused on topics of access to care, beliefs about genetics testing, experiences with healthcare systems, and reflections on how others in their ethnic communities may feel. For Hispanic/Latinx and African American participants, we also directly asked about ways to improve enrollment for those racial groups. Interviews were between 30 minutes to 1.5 hours long. All interviews were conducted on video conferencing software (Zoom Video Communications, Inc, San Jose, CA, USA) in English and transcribed interviews verbatim. Live Spanish interpretation by a certified Spanish medical interpreter was available for respondents who chose to converse in Spanish, though all patients in this set of interviews opted to speak in English.

Analysis

We present survey information as percentages; we did not make statistical inferences with respondent demographic information due to small sample size. For interviews, audio recordings were transcribed and analyzed for themes in NVivo qualitative data analysis software, version 1.7.1 (QSR International, Burlington, MA, USA). As this was an exploratory analysis of experiences, we took a grounded-theory approach and codes were generated inductively from responses. Codes were applied to whole excerpts and co-occurrences were permitted. After review of all available transcripts, codes were collected into themes by the medical student researcher after adjudication with faculty advisor SAL.

Ethics statement

This qualitative study was approved by the Yale University Institutional Review Board (ID# 2000032894). Informed consent was obtained from all respondents prior to the questionnaire, which also included the interview. The medical student investigator worked independently from the PGDP care team.

Student contributions

The medical student created the questionnaire and interview guide with conceptual input from faculty advisor SAL and assistance from NH for study design and review. The medical student researcher directly recruited participants by email, interviewed all respondents, coded all transcripts, and employed statistical and qualitative analysis of surveys and interviews to derive the study's key conclusions.

Results

Questionnaire

We invited 175 participants to the initial questionnaire and 36 participants completed the survey (**Table 2**). Of the 36 respondents, 29 noted moderate or significant barriers (**Table 3**). Of all the factors, knowledge about the PGDP program was the most commonly cited. Not knowing the program existed (42%) and Not knowing if they would qualify for the program (36%) were the most frequently noted as moderate/major barriers. When asked to rank the top three barriers, Not knowing the program existed also had the most number of

	Survey respondents
Total	36
Female	94.4%
Relation to patient	
Parent of patient	91.7%
Patient	8.3%
Race	
Hispanic/Latinx	11.1%
White non-Hispanic/Latinx	83.3%
Black non-Hispanic/Latinx	0%
Asian non-Hispanic/Latinx	5.6%
Other non-Hispanic/Latinx	0%
Primary language at home	
English	100%
Marital status	
Not married, living alone	5.6%
Not married, cohabiting	8.3%
Married	83.3%
Divorced	2.8%
Education (highest level attained)	
Secondary school	27.8%
College/University (Undergraduate degree)	30.6%
College/University (Graduate degree)	41.7%

Table 2. Demographics of survey respondents.

Question	Not a barrier or Minor barrier	Moderate or Major barrier
My healthcare providers or I did not know this program existed.	58%	42%
I did not suspect the medical problems could be caused by a genetic condition.	78%	22%
I had trouble communicating with healthcare providers because we spoke different languages.	97%	3%
I had a hard time scheduling appointments with my primary doctor or specialists.	72%	28%
I knew about the program but my healthcare providers or I did not think our medical condition would qualify for this program and/or wanted to wait for other diagnostic tests first.	64%	36%
I knew about the program, but I had concerns about genetic testing.	100%	0%
I knew about the program, but I did not trust the program.	94%	6%
I knew about the program but I was not sure how to join.	86%	14%
I knew about the program but I was concerned with how much this program would cost.	86%	14%
I knew about the program, but I did not think the program would be helpful to us.	94%	6%
I knew about the program, but other family members did not want to participate.	94%	6%
I had difficulty arranging for time off work to come here.	94%	6%
I had difficulty arranging for childcare to come here.	89%	11%
I had difficulty arranging for transportation to come here.	100%	0%
I was worried that participation would cause legal issues for me or my family.	100%	0%
Other	78%	22%

Table 3. Responses to the survey asking patients or family members to estimate the most significant barriers to reaching PGDP.

votes (13), with uncertainty about qualifying for the program (10), difficulty scheduling appointments with specialists (8), and not suspecting a medical condition (7) also noted as top barriers (**Fig. 2**). For those who responded in the Other category (22%), two described poor communication by the PGDP, two were delayed by genetic

samples being lost in transit, two felt a “lack of urgency” by medical doctors prior to PGDP, and one was concerned about a risk for life or health insurance discrimination by participating in genetic research.

By comparison, Family buy-in, Financial cost, Arranging childcare, Transportation, Language barriers, Concerns about genetic testing were not significant barriers for almost all respondents. Seven (19%) respondents listed no moderate or major barriers. Of these, two were direct inpatient referrals from PGDP, four were

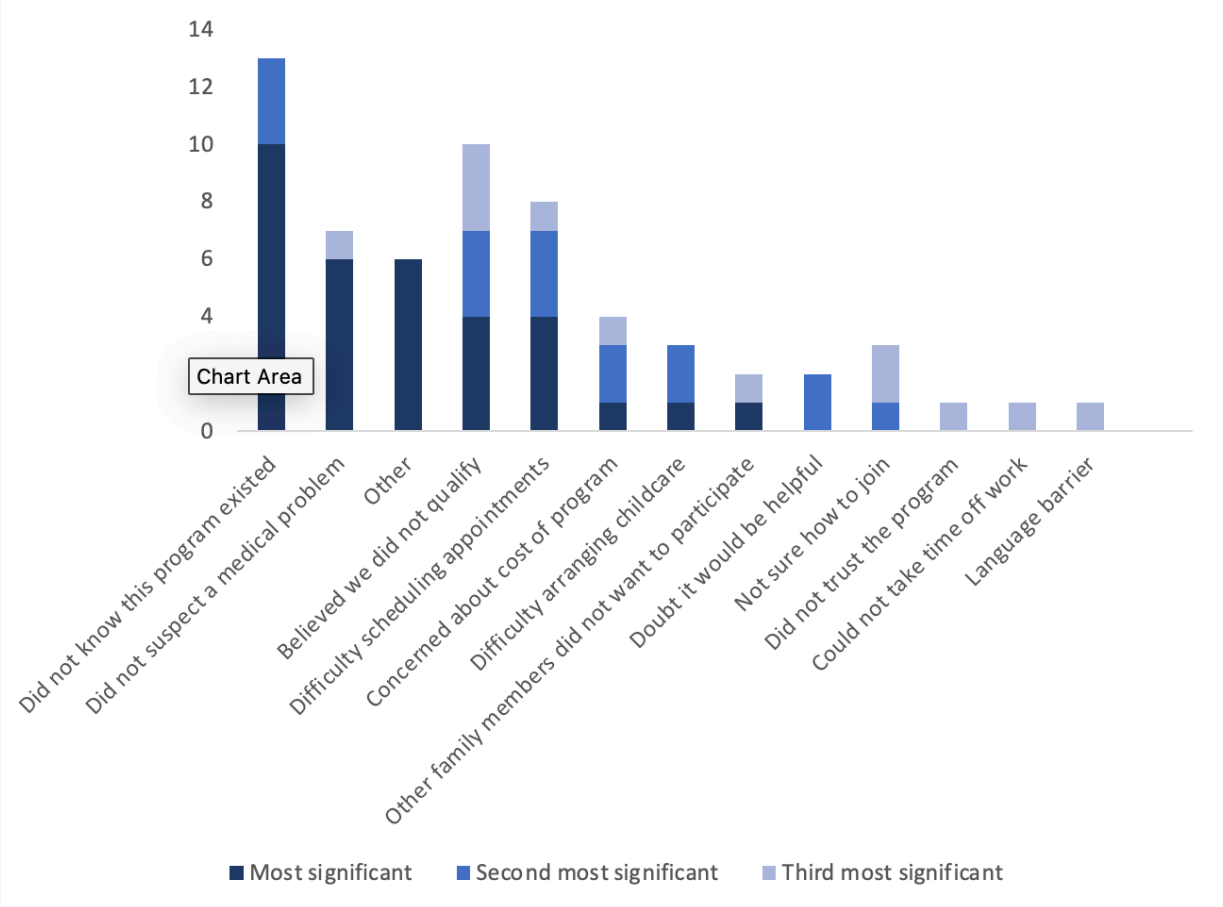

Fig. 2. Comparison of the importance of each barrier along the diagnostic odyssey. Respondents were asked about the significance of sixteen barriers they may have encountered on the diagnostic odyssey. Moderate and Major barriers were re-presented to participants, who were asked to rank what they considered their top three barriers.

referrals from the Yale Genetics program, and one was referred by another Yale specialist.

Semi-structured interviews

33 questionnaire respondents were interested in participating in follow-up interviews; eleven accepted the interview invitations (**Table 4**). One participant was the patient while the others were the parents of patients. Thematic analysis revealed three main themes (**Table 5**), detailed below: 1) experiences of the diagnostic odyssey, 2) tools used to navigate uncertainty, and 3) direct experiences and perceptions of the PGDP.

	Enrolled (11 of 30)	Not Enrolled (0 of 30)
Hispanic, any race	2/10	0/10
Black non-Hispanic	0/10	0/10
All other races, non-Hispanic	9/10	0/10
White	8	
Asian	1	

Table 4. Race/ethnicity of participants enrolled for in-depth interviews.

Theme 1: Challenges along the diagnostic odyssey

Respondents described the diagnostic odyssey not just through the process of finding a diagnosis, but the accumulation of one unanticipated new medical problem after another without any clear causes or treatments. Eventually, this led to

Experiences of the diagnostic odyssey	Tools to navigate uncertainty	Perceptions of the PGDP
• One problem after another • Accumulate multiple specialists • Resource-intensive, slow, and emotionally draining • Progress based on luck • Lack guidance from professionals • Insensitive or dismissive providers	• Full-time job • Leverage personal and professional connections • Social groups • Seeking multiple opinions • Medical knowledge • Experience with healthcare system	Barriers and reservations: • Yet another medical appointment • Mistrust of research • Traumatized by past experiences • Diagnosis anxiety Motivations: • Move forward • Family planning • Access treatments and services

Table 5. Summary of themes from the semi-structured interviews.

respondents going to ever-expanding lists of medical appointments with multiple specialists.

> He ended up with asthma at 18 months-ish. And then so we were referred to respiratory. And then just after turning two, he had a kidney stone so we saw nephrology… And then, finally, after seeing kidney, they sent us to GI because he was having issues. And then from GI, we ended up with endocrinology after that… So it's been one specialist after another, after another, after another… No doctor would take the blame as to what was going on.

Respondents described that the diagnostic odyssey was resource-intensive, slow, and emotionally draining. They felt that access to specialists was mostly through serendipity rather than sincere efforts from their doctors, and several were frustrated by the lack of guidance from their doctors. Respondents wanted doctors to think more holistically about their conditions and offer guidance as to what the “next steps” for diagnostic work-ups or treatments could be.

> GI said that her GI symptoms weren't something that he could treat… and kind of dismissed us from his practice even though she had recurrent diarrhea and stomach pain. And he said it was more of a rheumatology problem, but we

didn't have anybody to refer us to rheumatology until this one ER doctor put the referral in.

Lastly, respondents described dismissive or insensitive comments made by medical professionals that created self-doubt and eroded relationships with doctors.

> I described what happened to me and my partner described, and he had had actually taped the aphasia, the seizure. And they were like, "Well, if you really had [a seizure], then you would've come here last night when it happened.
>
> - Patient

> One provider took one look at my daughter and said, “Well, she'll never be the captain of the soccer team.” Yeah, she was having trouble walking at the time. But we don't need to hear that.
>
> - Parent

Theme 2: Tools to navigate the uncertainty

Respondents described a number of methods they used to navigate the diagnostic odyssey. The first was that whenever possible, one parent had to treat the diagnostic odyssey as a full-time job to further their diagnostic journey by keeping track of medical information, researching the next tests and experts, and coordinating care.

> [When the patient was alive] I could just focus on advocating for her, bring her to therapy, bringing her to appointments. I mean, that was our whole life.
>
> - Parent

> I had binders. So when they would say something, I'd go home and research it, print up a little thing, put it in the binder. And if it came back, “Hey, it's not that,” then I'd just take it and put it in the back of the binder just because hey, that was an option. And then I had notebooks and just constant research and always just being on my toes.
>
> - Parent

Respondents leveraged personal or professional connections to access appointments or services.

> My husband was working at Yale at the time as a physician, and we actually had to talk to his division chair to say, “Please, is there any way that you can get [a appointment for genetics] ...,” which we feel bad about doing. Obviously, everyone hates to do that kind of thing, but we were talking about not being able to get in for a year. We couldn't even talk to [genetics]. And eventually, his department helped us. And it was still four or five months before we could get in with them.
>
> - Parent

Similarly, parents tapped into in-person or social media groups to gain advice from other parents in similar situations.

> Through a parent group that I was a member of on Facebook, someone told me about [condition]. So, I don't know I just started digging and digging, and then one thing led to another and this woman that I was chatting with said, “Oh, well, maybe it's an immune issue. I see this great doctor in [Location]. Her name is Dr. [Doctor's name] and she's a mast cell activation specialist and, you know, Ehlers-Danlos and that's her niche.”
>
> - Parent

Patients commonly visited different doctors in the same specialty, or doctors in different healthcare systems altogether for different perspectives, or when they were “not getting the answers.” Two respondents described enrollment in more than one advanced genomic diagnostic program. Several respondents felt they had high medical literacy or familiarity with navigating the healthcare system.

> [I have a] PhD in [liberal arts]… my husband is a doctor… we thought that we'd be able to understand it, and I totally don't understand any of the stuff that [PGDP is] doing.
>
> - Parent

Theme 3: Perceptions of the PGDP

Subtheme: Barriers and reservations

Nearly all respondents mentioned a major barrier to reach PGDP was that it was not widely advertised or known to other doctors. Respondents thought that PGDP should make concerted efforts to advertise the program: social media, pediatricians,

geneticists, developmental services, special education teachers, rare disease groups. When they did hear about the program, respondents were unclear about how the program worked. Sometimes, this meant not understanding how the program would be different from a regular doctor's visit.

> [I thought] "Oh, no, another doctor we got to add to our thing, or another appointment we got to add when we already have so many appointments."
>
> - Parent

The diagnostic workup itself can be traumatizing for children and parents.

> I almost pass out every single time my younger daughter has to get a blood draw because she has such an aversion to needles now, after going through so many blood draws, especially in a short amount of time when she was little, that it becomes a really traumatic experience for us.
>
> - Parent

Respondents cited mistrust of the program as another cause for hesitation, including suspicions around what PGDP planned to do with the genetic information, how much it would cost parents, and the future privacy concerns. One single respondent also mentioned a major hesitation with anxiety facing a potential diagnosis: "it's almost better to not know anything than to know that you have something unknown."

Subtheme: Motivations

Respondents described a wide range of specific motivations for finding a diagnosis and participating in the PGDP. Several described interests related to scientific discovery and personal curiosity, but families of children in particular wanted answers so that they could "move forward." Parents wanted to know if they were at

fault for the child's medical problems, what will happen to their child in the future, and whether these problems would impact any other family members or future children.

> Before you guys found his mutation, I was blaming myself. I assumed I'd done something to cause his mutation.
>
> - Parent

> Our daughter was young and she had already had a seizure, and we really needed to figure out what was going on with her, especially for family planning. So if you want more kids, you can't afford to wait five or 10 years to get all the answers.
>
> - Parent

Parents similarly hoped that a diagnosis could lead to medical treatments such as novel drug therapies or could help open "a whole other door for services."

Finally, when reflecting on their experiences with PGDP, respondents expressed gratitude for the program's services. The program allowed respondents to access new services or treatments, and helped families achieve closure.

> It's been a long three years, but after [PGDP diagnosis], there's a targeted therapy that [PGDP found], which made a significant difference in my daughter's life… it stopped her fevers, it stopped the bleeding issues, it stopped how severe she gets when she does get sick. And without the program, I don't think that we would be in the same position we're in today.
>
> - Parent

> That was a huge relief for me and I really stopped blaming myself for it, so psychologically has a huge impact.
>
> - Parent

Discussion

Although several qualitative studies have highlighted the challenges of the diagnostic odyssey,[17,36] we are the first to our knowledge to ask families to objectively

quantify and compare specific barriers along their journeys. We found that the two most significant barriers to reach the PGPD are related to awareness about the program: realizing that such a program exists, and that their conditions would qualify for such a program.

Let us consider the first barrier: program obscurity. In interviews, participants elaborated that the obscurity of the program came from a lack of guidance on how to move forward in the diagnostic odyssey in general. A characteristic of the diagnostic odyssey is the accumulation numerous specialists over the course of the journey, but the care is often uncoordinated between providers.[33] Treatments and diagnostics are considered in isolation without an overarching strategy for what comes next after negative test results. Thus, instead of predictably escalating to a genomics program like PGDP, referrals are dependent on the specific provider's interpretation of clinical data, and awareness of such programs.

The second barrier is a *perception* of program exclusivity. In other advanced genomics programs, patients are required to submit applications with prerequisite testing, letters from physicians, the patient's signs and symptoms, and a personal narrative.[26] Indeed, the majority of participants interested in such programs perceive them as the last resort for diagnosis.[18] In our study, one-third of participants were delayed because they were similarly concerned that they would not qualify for the PGPD. In reality, PGDP is free and there are no pre-requisite requirements to participate.[27] We are currently in the process of reviewing our recruitment materials to directly address questions of cost, eligibility, and risks and benefits of participating.

Parents play an integral role in steering the diagnostic odyssey, often diverting significant resources and acting as the “care-captains” for their children when they lack the support of formal guides.[37] In this role, they regularly negotiate with healthcare providers, conduct their own research on potential programs or specialists, document extensively, and become experts on the signs, symptoms, and treatments for their children. We additionally view parental leadership as a *requirement* to successfully navigate the systemic barriers of the diagnostic odyssey. In our study, parents described learning on their own to navigate uncertainty, persistently looking for specialist referrals and diagnostic tests, all while dedicating themselves as the full-time caretaker for their child with complex needs. Individually, these are noble efforts. Collectively, they highlight the systemic barrier from the diagnostic odyssey that requires families to divert substantial financial resources, caretaker time, and active attention to move forward. All the other tools we described - seeking multiple opinions, leveraging personal connections, and planning the next steps - are mediated by the parent. If the parent cannot be the care-captain, then this could easily result in delays or even failure of the diagnostic odyssey.

We additionally found that parents rely on the guidance of peers to plan their next moves. Our respondents noted the importance of social connections as integral to the success of a diagnostic odyssey, and some explicitly asked to have a social community of families involved with PGDP. Undiagnosed communities, like communities for individual rare diseases, are an important source of information, emotional support, and motivation to further their journeys.[17,34,36,38] Parents may feel they struggle to find the “right” communities on social media that reflect their

undiagnosed conditions.[34] But in the melting pot of different conditions, participants in rare and undiagnosed disease groups are nonetheless able to share valuable, generalizable advice in symptom management or acquiring services.[34]

Social groups also give parents hope by seeing the progress for other families.[38,39] PGDP has an initial diagnostic yield of approximately 30%,[27] and for those who remain undiagnosed, reanalyses over months to years incrementally increase the likelihood of finding a diagnosis.[40] For struggling parents, seeing not just a diagnosis, but a diagnosis that leads to treatment changes, expectation setting, and family planning,[10,27] can greatly improve feelings of anxiety and depression[35] and remind them that AGDRPs are more than "yet another medical appointment."

Although we gathered data from 20% of the cohort, the majority of our survey participants were White non-Hispanic, well-educated, married parents living in the state of Connecticut. In this first round of data collection, our study had a low response rate of 11% from URMs. A lower participation rate of URM in surveys about medical experiences is not unusual and qualitative studies by other genomics programs report Hispanic and African American representation between 6.9% to 28%.[18,35,39,41]

Still, having a diverse set of perspectives is important in our research because some groups may be disadvantaged by specific barriers more than others. We used quotas, bilingual materials, monetary incentives, and direct communication to encourage URM participation, but these incentives did not result in a participation proportionate to PGDP cohort demographic. We will need additional strategies to increase engagement of under-represented minorities in the next iteration of this study.

Based on a toolkit developed to improve representation in research, we are reviewing our recruitment strategy and looking for ways to meaningfully involve more URM participants in our research.[42] We are expanding our recruitment medium to include phone calls, text messaging, physical mail, posters, and online advertisements. Based on participant feedback, we will be reaching out to rare disease social media groups, special needs schools, and government departments to find additional undiagnosed children.[42]

Conclusions

The most significant barriers to reaching PGDP were knowledge about the program's existence, recognition of an undiagnosed genetic disease, access to specialists, and fragmented care. Parents had the dual responsibility of caring for the substantial health needs of their children and leading the pursuit of diagnostic evaluation. Parents relied heavily on social groups, the internet, and personal connections to overcome a lack of direction and uncooperative providers. Specific motivators for PGDP participation included seeking closure, family planning, and having a formal diagnosis to access social services or treatments. We still need the perspectives of URM families along the diagnostic odyssey; efforts are being made for a more diverse recruitment for interviews in the future.

STUDY AIM #2: INVESTIGATE WHETHER SYSTEMIC BARRIERS EXIST IN HOW PGDP RECRUITS PATIENTS

Introduction

The qualitative data from Aim 1 helped us better understand the diagnostic odyssey for patients and families reaching the PGDP. In this Aim, we sought to characterize the demographics of PGDP participants, specifically highlighting three distinct routes that we thought had different levels of access to PGDP. These are: 1) direct enrollment by a PGDP team member in the pediatric intensive care unit (referred to as PGDP-Direct); 2) referral from Yale Genetics after a comprehensive, negative genetic work-up (referred to as Yale Genetics); and 3) a parent-as-care-captain journey through self-referral or by stumbling upon a provider familiar with PGDP along their journey (referred to as Other Referrals).

Our rationale requires some explanation. There is a robust PGDP representation among pediatric critical care doctors within the Yale-New Haven Health System (YNHHS) as the Director and Clinical Director of PGDP are both clinically active pediatric critical care physicians.[27] As such, PGDP clinicians can easily recognize and meet with families of critically ill children while they are on service, and directly invite these families to participate in the program.[10] This route is potentially the most straightforward, because it is not uncommon for the explanation of the program, consent, evaluation, and genetic testing to all be done in the span of a single inpatient

hospitalization. At the same time, the route is limited because it implicitly necessitates that the patient is hospitalized, and even more, sick enough to require ICU-level care.

The second is what could be considered as the “classic” journey where families or clinicians suspect a genetic disorder, obtain a referral to an outpatient clinical geneticist, and receive a formal genetic work-up.[30] Geneticists at our institution regularly refer patients to PGDP if they have negative or inconclusive results on ES. This route is characterized by extensive genetic testing and guidance along the diagnostic odyssey by genetic specialists, but presupposes access to a geneticist.

The last category of patients are those who find PGDP on their own or through serendipitous encounters with medical providers familiar with PGDP. In this journey, patients have little to no guidance on what or where their diagnostic pursuit will take them. Patients from this route may have extensive testing prior to referral, but may also have little to no testing. This may be the hardest route, because it requires looking for a program that they do not yet know exists, and therefore may require substantial resources and understanding of the medical system.

With this rationale, we hypothesized that these three journeys are not equally accessible to resource-limited families, and those who have been historically marginalized by the healthcare system.

Methods

Study population

Our study population is the same as what was described in Aim 1. Since we were looking at barriers in the U.S, we excluded all patients with international or unknown addresses. We included every identifiable PGDP participant that had race/ethnic, zip code, and enrollment status data.

Data collection

We used self-reported demographic information from an internal, HIPAA-compliant database. We used publicly available datasets from 5-Year American Community Survey (ACS), 2015 – 2019 for demographic comparisons with children in Connecticut,[43] and the 2016 – 2020 5-year ACS Public Use Microdata Samples (PUMS) dataset to estimate race- and state-adjusted incomes.[44]

Estimating income data

We estimated household income data of participants by cross-reference of zip codes against the American median household income survey. We chose this method due to two problems with simply comparing median household income data between races and within races across states. First, it is well-established that income is associated with race and ethnicity, with Black and Hispanic families consistently having the lowest household incomes.[45] Therefore, direct comparisons median household incomes of PGDP participants may reflect the existing disparities among races rather than biases specific to the diagnostic odyssey. Second, direct comparisons of in-state and out-of-state participants would have been similarly skewed by the differences in income across states, because Connecticut had one of

the highest median household incomes.[46] In other words, an out-of-state participant may have a relatively high income relative to their state and cost-of-living standards, but still have a lower income than the average participant within Connecticut.

Therefore, we used the PUMS dataset to create a normalized range of median household incomes into percentiles for each race within each state. Next, we calculated the median household income PGDP participant based on their documented zip code. Finally, we calculated the income percentile for the participant, relative to the participant's state of residence and race. We called this the income percentile. This method allowed us to compare how well-resourced a participant is relative to other families in their home state, and similarly, how well-resourced a participant is to their own race group.

Demographic comparison with UDN and the broader U.S. demographic

We lastly sought to compare PGDP demographics against external cohorts. We used the demographic data from a UDN article published in 2018 because it offered the most comprehensive demographics on both pediatric and adult patients in all sites.[29] UDN did not distinguish patients into White Hispanic and White non-Hispanic, so be conservative, we assumed that White race included all Hispanic patients and separated the proportion of Hispanic patients from White race. We compared the two programs to the U.S. Census data described earlier.[43]

Analysis

Pearson chi-square was used to compare categorical demographic data, summarized as percentages: race and ethnicity, enrollment status, and state of residence. Kruskal-Wallis one-way analysis of variance with pairwise comparisons was used to compare median household incomes (unadjusted and income percentile); Asian and Other races were excluded due to low sample size and heterogeneity.[47] A multinomial logistic regression model was used to estimate the odds-ratios for enrollment by factors of race, place of residence, and the route of referral. Income measures were excluded from the model because of its strong collinearity with race.[47] Significance was defined as p-value <0.05 after Bonferroni's correction for multiple comparisons.

Ethics statement

The study was approved by the Yale IRB (HIC: 1411014977) and written informed consent was obtained from participating families prior to data entry into the PGDP program database.

Student contributions

The medical student researcher conceived and designed this study as principal investigator, with conceptual input from faculty advisors SAL and LJ. The student investigator used publicly available datasets for county-level demographic information of children as well as statewide racially-adjusted household income levels.

Results: PGDP cohort demographics

Our cohort consisted of 252 patients and 80% of these had completed enrollment in the PGDP (**Table 6**). Internal referrals made by clinician-scientists directly affiliated with the PGDP accounted for one-half of all Hispanic/Latinx participants in the program, whereas white non-Hispanic patients made up the majority of non-PGDP referrals. Additionally, PGDP-Direct referrals had the highest percentage of Blacks and Hispanics (52%), compared to Yale Genetics referrals (27%), and Other Referrals (16%). Income was not significantly different among the three groups, both in unadjusted and adjusted terms.

	PGDP-Direct	Yale Genetics	Other Referrals (Includes self-referral)
Total	68	98	86
LIVE IN CONNECTICUT	91%	96%	43%
INCOME			
Unadjusted	$66,959 ($50,621 - $95,504)	$81,358 ($58,889 - $110,111)	$73,093 ($55,357 - $96,798)
Percentiles adjusted for state and race (percentile)	43 (27, 58)	51 (40, 64)	51 (40, 62)
RACE			
Hispanic/Latinx	43%	16%	11%
White non-Hispanic/Latinx	40%	57%	80%
Black non-Hispanic/Latinx	9%	11%	5%
Asian non-Hispanic/Latinx	4%	7%	4%
Other non-Hispanic/Latinx	4%	8%	0%
Enrolled in PGDP	66%	81%	88%

Table 6. **Demographics associated with three referral routes to PGDP**. PGDP-affiliated pediatric critical care physicians directly enrolled patients while they were hospitalized (PGDP-Direct). Patients followed by Yale Genetics were typically referred to PGDP if they had a negative genetic work-up (Yale Genetics). In Other Referrals, patients could be referred from individual providers, if they knew about PGDP. Additionally, there were no restrictions on patients referring themselves to the program.

In multinomial logistic regression, out-of-state participants were 5.2x more likely to be enrolled compared to In-state (**Table 7**). In contrast, method of referral and race/ethnicity were not significant predictors for enrollment.

	In State	Out of State
Total	193	59
INCOME		
Unadjusted	$76,970 ($55,033 - $101,806)	$72,075 ($55,019 - $95,367)
Percentiles adjusted for state and race (Percentile)	49 (34, 60)	51 (41, 61)
RACE		
Hispanic/Latinx	26%	10%
White non-Hispanic/Latinx	56%	75%
Black non-Hispanic/Latinx	8%	10%
Asian non-Hispanic/Latinx	6%	3%
Other non-Hispanic/Latinx	5%	2%
Enrolled in PGDP	75%	95%

Table 7. PGDP participants, compared by home state of residence. *In State,* state of Connecticut.

In terms of unadjusted income, in state White non-Hispanic (IS-White) had significantly higher income compared to in state Black non-Hispanic (IS-Black), in state Hispanic (IS-Hispanic), and out of state White non-Hispanic (OS-White) ($p < 0.001$, $p<0.001$, and $p=0.04$, respectively) (**Fig. 3**). However, when adjusted for the income levels of the racial and ethnic communities in the participant home states, median household incomes for out of state participants rose. As a result, IS-White no longer significantly differed from OS-White and OS-White became significantly higher than IS-Black and IS-Hispanic. Notably, out of state Black non-Hispanic (OS-Black) participants had significantly higher income percentile compared to IS-Black.

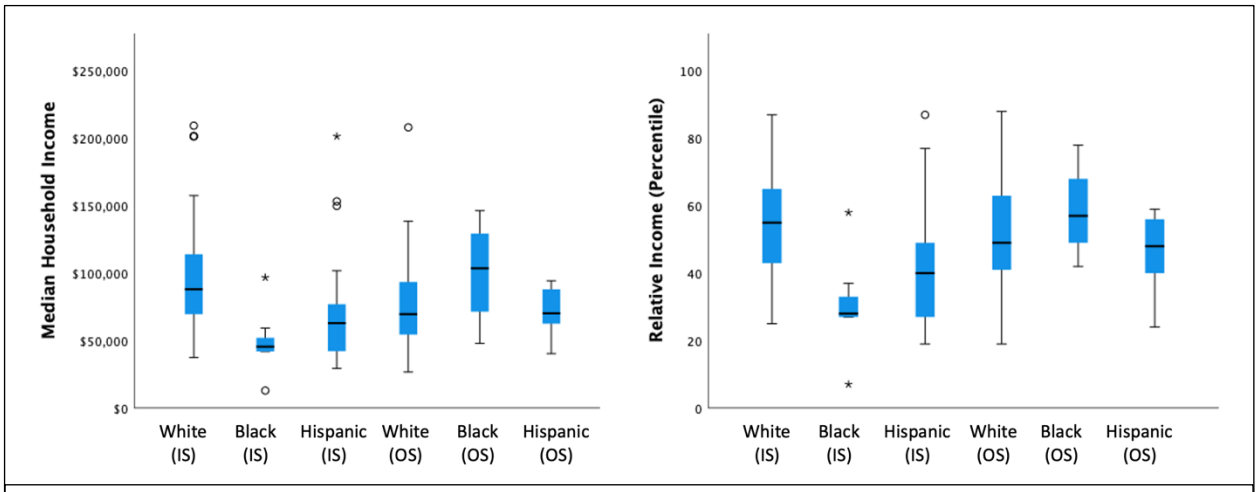

Figure 3. Unadjusted and adjusted income levels across White, Black, and Hispanic participants of PGDP. Median household income was extrapolated from ZIP codes. IS, Live in state. OS, Live out of state.

Lastly, we compared the demographics of the three referral routes for PGDP, published demographics from the UDN pediatric cohort,[29] and all U.S. children (**Table 8**).[43] We made five statistical comparisons: four between the advanced genomic diagnostic programs with children in the U.S., and one between PGDP-Direct and

UDN. After correction for multiple comparisons, PGDP-Direct, PGDP-Other Referral, and UDN-Accepted Pediatric were significantly different from the racial proportions of children in the U.S. ($p=0.008$, $p<0.0001$, and $p=0.003$), while PGDP-Yale Genetics was not (0.638). PGDP-Direct and UDN were significantly different ($p<0.0001$) with PGDP-Direct having a higher proportion of URM (52%) compared to UDN (23%).

Race/Ethnicity	PGDP-Direct	PGDP - Yale Genetics	PGDP - Other Referrals	UDN (Accepted pediatric applicants)	U.S. Census (children)
Hispanic	43%	16%	11%	18%	25%
White non-Hispanic	40%	57%	80%	56%	50%
Black non-Hispanic	9%	11%	5%	5%	14%
Asian non-Hispanic	4%	7%	4%	7%	5%
Other non-Hispanic	4%	9%	0%	14%	6%

Table 8. **Comparison of PGDP demographics with published data of the Undiagnosed Diseases Network, a peer genomic diagnostic program, and the U.S. as a whole**. All except PGDP-Yale Genetics ($p=0.638$) were significantly different from the U.S. Census. PGDP-Direct was significantly different from UDN.

Discussion

We found that the racial/ethnic composition of patients with undiagnosed diseases differed greatly depending on how they were referred to PGDP. The majority of patients live within the state of Connecticut, which indicate a geographical bias in our population of undiagnosed children. These trends all point to disproportionate attrition of under-served minorities as the number of barriers to health services increase, such as family support, logistics, and financial barriers.[30]

Why might there be a more equitable representation of under-served minorities from direct recruitment by PGDP-internal providers? Based on the results of Aim 1, we know that continuity of care – the ease that patients can progress in a healthcare system – is missing in the diagnostic odyssey and is suggested as a top priority by our participants. Across the three pathways to reaching PGDP, direct recruitment theoretically has the fewest barriers to reaching the program, because the referral process follows a bedside-to-bench-to-bedside philosophy. PGDP-associated providers are uniquely positioned to treat a representative sample of children who would most benefit from the program.[10] PICU admission does not require prerequisite criteria, appointments, waiting, or financial resources. Critical care pediatricians see a wide range of undifferentiated syndromes in critically ill children, and have a higher proportion of these children with rare, complex genetic conditions.[10] While there always a concern of individual provider biases for different racial/ethnic groups, critically ill children raise red flag symptoms and these providers have a low threshold for investigating unsolved medical problems. In contrast, genomic diagnostic programs tied to Genetics clinics skew disproportionately to White, non-Hispanic, more educated, higher income, and privately insured families, even when the local community is ethnically diverse.[15]

If PGDP-affiliated providers are the best positioned to find undiagnosed children, why then do PGDP-internal referrals have the lowest rate of enrollment completion? Several factors likely contribute. After adjusting for inherent income differences across states and race groups, we found that the OS race groups had similar or statistically higher relative financial resources compared to IS groups.

Families who make it to the program through Yale Genetics and Other Referrals are self-selective for those actively looking for a diagnosis, and we further suspect that these are “motivated” patients/families with resources. These families, especially ones that have the resources to travel to Connecticut to seek out PGDP services, would explain why living outside the state of Connecticut carries a 95% enrollment completion rate.

In contrast, PGDP-Direct referrals carry a far lower enrollment completion rate. We do not have a definitive answer for why this is the case, so we will offer some speculations. PGDP-Direct may be finding patients early in their diagnostic odyssey, and a portion of these patients will decline to continue. Alternatively, PGDP-Direct may be reach patients at a very late stage in the diagnostic odyssey – to the extent that families have already given up. Parents might feel that a diagnosis would not be able to change their circumstances anyway. Finally, it is plausible that in the ICU setting, families are willing to participate if the program can help their child, but may question its utility on follow-up appointments after they are discharged from the hospital. Further investigation is required (perhaps through future patient interviews) for this unexplained observation.

In our study cohort, the proportion of Hispanic and Black non-Hispanic patients progressively lowered as the degrees of separation increased between the referral route and the program. Both UDN and PGDP-Direct had significantly different demographic compositions relative to the general public, but these were in different directions. While UDN had a lower URM composition (23%) compared to the broader U.S. (39%), PGDP-Direct was higher. The local community around PGDP - New Haven

County - has a higher proportion of Hispanic (31%) and African American (14%) children compared to the broader U.S.[44] Since most PGDP-Direct patients live within the state, it is plausible that the higher proportion of URM in the local community is being represented in PGDP-Direct enrollment.

Conclusions

Racial and ethnic demographics of PGDP, and more specifically the proportion of URM, changed substantially as the referral route to PGDP became comparatively less direct. Residence outside Connecticut was a strong predictor of completed enrollment, possibly through a combination of resources and motivation to join the program. Direct, inpatient PGDP referrals have a lower rate of completed enrollment, but a higher representation of URM compared to a peer genomics diagnostic program and the general U.S. population.

STUDY AIM #3: FIND CHILDREN WITH POTENTIALLY UNDIAGNOSED DISEASES IN THE YALE-NEW HAVEN HEALTH SYSTEM

Introduction

A major challenge of conducting outcomes research on undiagnosed children is that the most precise sources of information on these patients – cohorts from advanced genomics diagnostic programs – only capture participants who have successfully navigated the diagnostic odyssey. As such, we do not see the patients that are not there. This problem is called survivorship bias: when the sample is not representative of the cohort because those who do not "survive" to the end are not considered.[48]

If we assume that all racial and ethnic groups have an equal chance of being born with an undiagnosed genetic disease, then where are the missing children? The existing literature on diagnostic program cohorts are of "survivors" who already overcame the other obstacles along the diagnostic odyssey, with demographic biases that have been a recurring theme within this thesis: families are more likely to be White, non-Hispanics with a higher education level than the general population.[26,29] Similarly (or perhaps consequentially), the published studies on experiences with undiagnosed diseases are predominantly from White non-Hispanic participants.[18,35,39]

From Aim 1, we established that families with undiagnosed diseases are often expected to carry the burden of continuing along the diagnostic odyssey. Those

lacking resources, education, contacts, or other key assets are unlikely to present to advanced genomics diagnostic programs. From Aim 2, we observed that the more distant the referral route is from the PGDP, the lower the proportion of URM we see, suggesting that significant care-captaining is required to compensate. Even PGDP-Direct referrals are limited; for example, children admitted to the general pediatric ward would not typically be noticed by the critical care physicians who are part of PGDP.

Therefore, a more systematic way of finding undiagnosed children is needed to broadly capture the diverse population of patients who could benefit from advancements in genomic medicine. In our final Aim, we designed a screening system that used discrete clinical criteria – namely, billing codes - to identify patients who could potentially benefit from the program.

Methods

We sought to develop a method to search the electronic medical record to find patients with potential for having undiagnosed genetic conditions by using International Classification of Disease, Tenth Revision, Clinical Modification (ICD-10-CM) diagnosis codes.

We began by manual review of all ICD-10-CM codes with two reviewers with medical knowledge (the student and faculty advisor), with the decision to use an individual code as part of the search made by consensus between the two reviewers; if a consensus was not reached, then the code was kept as part of the search (not removed).

Broadly speaking, diagnostic codes that had little to no relevance to genetic diseases, such as those related to trauma, infections, acquired conditions or even known genetic diagnoses were excluded, as were non-diagnostic codes (such as hospital disposition) (**Table 9**). In order to maximize sensitivity, a patient was flagged given the presence of one or more ICD-10-CM code that the reviewers felt could indicate a undiagnosed genetic condition; but the presence of a ICD-10-CM code not on the list would not exclude the patient. Our method resulted in a list of 374 unique ICD-10-CM codes associated with clinically descriptive syndromes, rare signs and symptoms, rare diseases, or undiagnosed diseases. These represented X unique top-level code categories and Y systems out of the Major Classifications of Disease [Add in data, verify numbers].

Excluded code types	Example
Traumatic injury	"Walked into lamppost, initial encounter" (W2202XA)
Typical infections in pediatric populations	Acute bronchiolitis due to Respiratory Syncytial Virus (J210)
Common clinical syndromes without literature of known genetic associations	Regular astigmatism (H522)
Diseases with clear environmental etiologies (i.e., drug-induced, toxins)	Drug or chemical induced diabetes mellitus with neurological complications with diabetic neuropathy, unspecified (E0940)
Descriptions of health encounters or patient characteristics	Problems related to medical facilities and other health care (Z75)

Table 9. Categories of excluded ICD-10-CM codes.

We then used these codes to search medical records. in the Epic Systems electronic health record (EHR) of the Yale-New Haven Health System (YNHHS) with assistance from the Joint Data Analytics Team (JDAT). We elected to use inpatient

records, theorizing that children requiring hospitalization would have the least barriers to accessing medical care and also given the high representation of rare diseases amongst hospitalized patients. Additionally, in Aim 2 we determined that our inpatient-based PGDP-associated physicians had the greatest enrollment of URM patients; here, we wished to compare those percentages with potential patients from the inpatient setting. Therefore, we searched in all pediatric inpatient hospitalizations (age <18 years at time of admission) from January 1, 2016 (when YNHHS started using ICD-10-CM codes) to December 31, 2021 for patients with the presence of one or more of the 374 codes selected on our screener.

Chart review of screened patients

We followed the screen of ICD-10-CM codes with a manual chart review by the student and faculty advisor of the first 10% of the resulting patients (by date of hospitalization) to determine those with a high suspicion for undiagnosed genetic disease. This chart review aimed to exclude those with known genetic diagnoses and those with low or no clinical suspicion for underlying genetic disease (**Table 11**). This group of the first 10% of patients was used to estimate the characteristics of the cohort as a whole and to compare racial, ethnic, and socioeconomic characteristics with known PGDP participants.

Inclusion criteria	Example
Rare clinical syndrome without environmental or genetic explanation	Ohtahara syndrome; West syndrome
Common, descriptive clinical syndrome with other systemic manifestations, without environmental causes	Autism spectrum disorder with recurrent seizures; congenital heart defects and scoliosis
Documented family history of similar, unexplained illnesses	Mother and two daughters have unexplained recurrent fever syndromes
Extended gene panel analysis with negative results	No candidate gene mutations revealed in genome sequencing
Ongoing genetic diagnostic evaluation with loss to follow-up	A family being followed by Yale Genetics but moved to another country before testing could be performed

Exclusion criteria	Example
Convincing environmental explanation	Cerebral palsy due to neonatal hypoxic brain injury
Diagnosis verified through genetic testing	Trisomy 21; Spinal Muscular Atrophy with genetic diagnosis
Syndrome with polygenetic etiology, in absence of other notable symptoms	Autism spectrum disorder, in absence of other concerning signs or symptoms

Table 10. Inclusion and exclusion criteria with examples. General criteria used during chart review of patients screened by diagnostic codes to determine whether the patient would be appropriate candidate for the PGDP

Analysis

We used the NumPy, PANDAS, and SciPy packages for the programming language Python for statistical analysis. Demographics and outcomes were compared between cohorts using Pearson Chi-square tests for categorical variables and two-sided Mann-Whitney U tests for continuous variables. Bonferroni correction for multiple comparisons were made when appropriate, and explicitly described when performed. Significance was defined as $p<0.05$ for all tests.

Ethics statement

This study was conducted in accordance with the ethical standards of the Yale University Investigation Committee and approved by the committee (HIC # 2000032153). Informed consent was waived as per federal regulation 45 CFR 46.116(d) due to minimal risk to the subjects. The student and contributors have no financial conflicts of interest.

Student contributions

The medical student researcher conceived and designed this study as principal investigator, with conceptual input from faculty advisors SAL and LJ and methodological guidance from PA. The student investigator developed the ICD-10-CM code list with SAL, submitted the database query to JDAT and was directly responsible for analysis and interpretation of the raw returned data and associated chart review.

Results: Epidemiology of undiagnosed children

With these codes we found 10,772 encounters with YNNHS representing 1,648 unique patients admitted during the study period. Of these, 36 patients were known PGDP participants. The racial/ethnic characteristics of the remaining 1,612 patients was not significantly different from the demographics of children in New Haven County[44] (*p=0.164*). We used the first 10% of patients as the sample cohort for manual chart review identified as 162 consecutive patients arranged by date and time of first hospital encounter. To assess representativeness of the sample, we compared the demographics of the sub-cohort with the full cohort (**Table 11**). The median age of the

sub-cohort was significantly older than the remaining group (10.8, p=0.0423) and more male (72.8%,p=0.0025) but not significantly different in race (p=0.7731).

	Whole cohort	First 10%	p-value
Total	1612	162	
Female	40.6%	27.2%	0.0025
Age at study endpoint or death	10.3 (5.6, 16.9)	10.8 (7.4, 16.0)	0.0423
Race/Ethnicity			
Hispanic/Latinx	33%	36%	
White non-Hispanic/Latinx	37%	32%	
Black non-Hispanic/Latinx	21%	24%	
Other non-Hispanic/Latinx	9%	8%	0.7731

Table 11. Comparison between the whole cohort and the first 10%.

Amongst inpatients undergoing chart review, 71 (43.8%) had medical histories with high clinical suspicion for an undiagnosed genetic disease, and 91 (56%) did not – either because they did not have medical histories consistent with undiagnosed conditions, or had already received a definitive genetic diagnosis (**Table 12**). The undiagnosed cohort had significantly more females (p<0.001). The median ages of the cohorts were not significantly (p=0.254), but demographics were (p<0.001). De-identified descriptions of the first 10 patients classified as undiagnosed are available in **Appendix C**.

The cohort of undiagnosed children did not differ significantly in age (p=0.063) or sex (p=0.053), but did have significantly different demographic composition (p<0.001) compared to the patients known to PGDP. Specifically, Undiagnosed had twice the proportion of Black non-Hispanic children, with a proportional decrease of

White non-Hispanic children relative to PGDP. Undiagnosed was significantly different from New Haven County demographics (p=0.033), while PGDP was not different from New Haven County (p=0.626) (**Fig. 4**).

	Undiagnosed	Not undiagnosed	PGDP
Total	71	91	36
Female	35.2%	20.9%	44.4%
Age at study endpoint or death	10.7 (6.9, 14.3)	11.0 (7.8, 16.6)	6.7 (5.6, 13.7)
Race/Ethnicity			
Hispanic/Latinx	31.0%	40.7%	27.8%
White non-Hispanic/Latinx	31.0%	31.9%	50.0%
Black non-Hispanic/Latinx	23.9%	23.1%	11.1%
Other non-Hispanic/Latinx	14.1%	4.4%	11.1%

Table 12. A demographic comparison between Undiagnosed, Not undiagnosed, and PGDP-enrolled patients.

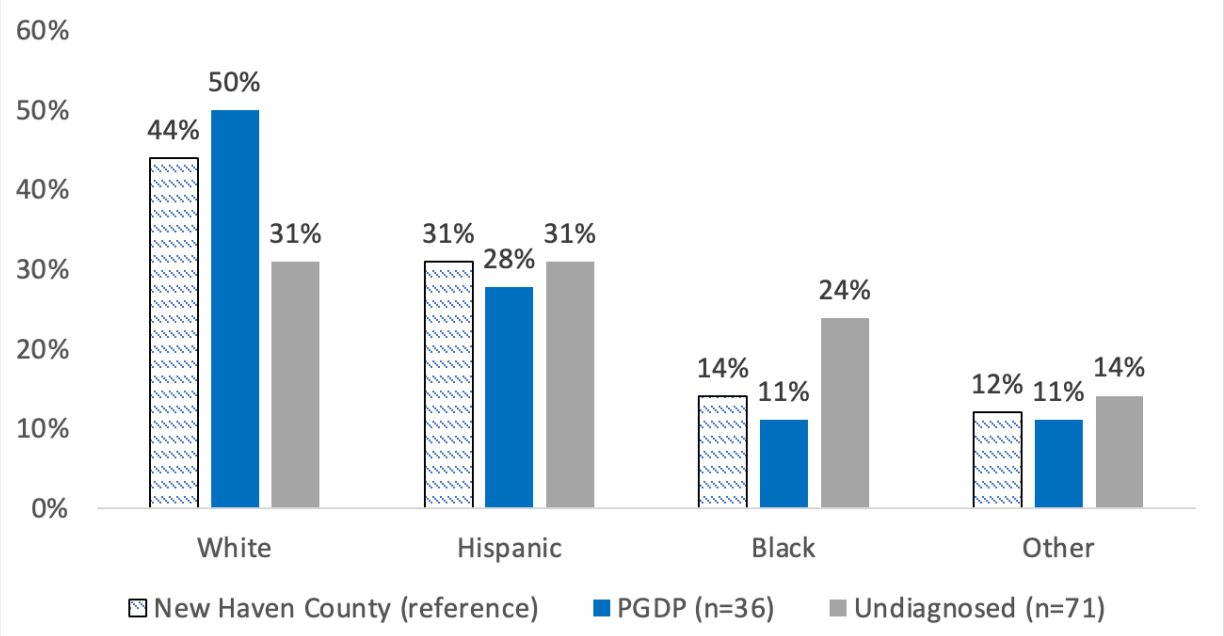

Figure 4. Racial comparison between PGDP, Undiagnosed patients, and New Haven County demographics.

Discussion

In this Aim, we created a reproducible system to identify patients with potential undiagnosed genetic conditions. Our strategy allowed us to leverage the greater URM representation on inpatient services, similar to what was seen in the PGDP-Direct referral route, without being limited by where patients were admitted and which providers were on service.

After chart review of the first 10% of children passing the ICD-10-CM screen, we found that approximately 44% had potentially undiagnosed genetic conditions. If we assume that the proportion continues for the remaining 90% of our screened cohort, then an estimated 700 hospitalized children over the last six years could have been referred to our program but were missed. Putting this number into perspective, during the same timeframe 36 from this group were actually referred to PGDP. Thus, for each patient referred to PGDP, there are perhaps 20 similar patients that were missed. Given that the total number of participants since PGDP's inception in 2015 – including the deidentified patients from collaborations with other programs – is over 700 (unpublished data), this search could possibly double the number of undiagnosed children in the program.

Our projections for undiagnosed children may appear large, but it is comparable to other similarly focused estimates. One single-center chart review of all emergency department visits found that approximately 18% of these visits were by patients who could have benefitted from additional genetic testing.[49] Similarly, a study of a de-identified database estimated the proportion of suspected or already diagnosed

genetic diseases to be approximately 9.4% of all unique pediatric patients and 45% of critically ill newborns.[23]

Although we did not have YNHHS-specific data for all pediatric admissions, other studies can provide a general estimate for our denominator. A study of 49 U.S. pediatric hospitals reported that 3,372,839 unique pediatric patients were hospitalized across 49 hospitals in a recent ten-year period.[50] That means on average, a pediatric hospital might see 6,883 unique pediatric patients in any given year. Over the six-year period of our study – with a denominator of 41,298 – the estimated 700 undiagnosed patients found in the EHR represent 1.7% of all unique patients, which is reasonable given that the quoted genetic disease prevalence of 9.4% in the previously mentioned study includes both suspected and confirmed genetic diseases. Importantly, this screen of inpatients only captured 36 out of 208 (17%) PGDP enrollees in the YNHHS EHR, suggesting that inpatient screens using ICD-10-CM codes alone, while helpful, may not be sufficient.

Notably, the Undiagnosed cohort is more demographically diverse compared to those known to the PGDP, and this observation might reflect a real unmet need for genetic diagnosis in URM. Genetic testing for URM have a lower diagnostic yield than for White non-Hispanic patients because the literature on pathogenic alleles for most genetic conditions have been focused on those of Caucasian ancestry.[31,51] But just as important, URM have decreased access to genetic testing, due to under-recognition of clinical syndromes by providers, the patient's resource limitations, and the culturally-influenced negative perceptions of genetic testing.[30] Combined, these factors should theoretically lead to a higher proportion of URM to be undiagnosed relative to White

non-Hispanic patients; this was consistent with our results, suggesting that these URMs are being marginalized in genomic medicine.

Our study has several points where the results could be affected by researcher subjectivity. We attempted to strengthen the code selection process and patient chart review with examples and guidelines, but ultimately, these processes require an element of clinical interpretation that may vary between different research groups. Moreover, although we tried to be broad in capturing potentially undiagnosed patients, we only captured one-sixth of the known PGDP cohort with this methodology. Our study likely underestimates the true prevalence of undiagnosed hospitalized children at YNHHS, and should be viewed as a methodological reference for future studies to refine and expand upon.

The use of ICD-10-CM codes as a screening tool is limited by a lack of more specific clinical data, possibly leading to false positives or negatives. Furthermore, simply grouping patients together as “Undiagnosed” (as we have done) can hide key distinguishing features between different diseases and presentations. We will continue to add clinical information to our dataset, with the hope of eventually computing predictors for enrollment in a similar way to Aim 2. These may include differences between individual providers, the numbers and types of departments that participate in the patient’s care, presence of genetic testing, and clinical phenotype. Furthermore, we intend to reach out to families in the Undiagnosed cohort directly, to invite them to our program and gauge the level of interest in the “missing” population. We will also invite these families to participate in our surveys and interviews from Aim 1 - our

Undiagnosed cohort presents a new set of potential perspectives to add to our current understanding of barriers to reaching PGDP.

OVERALL RESULTS AND CONCLUSIONS

Through interviews and surveys with PGDP participants, significant barriers to reaching our advanced genetic diagnostic program included a lack of guidance for the diagnostic odyssey and not even knowing that such programs exist. Important tools for families to advance the odyssey were active participation by parents as care-captains and guidance and support from others via social media. Analysis of the three primary referral routes to PGDP revealed different racial/ethnic demographics, with URM over-represented in direct inpatient enrollments but under-represented with greater distance to the program. Designing a broad ICD-10-CM-based screening strategy to find the missing undiagnosed children through the EHR allowed us to find 1,648 previously unknown, undiagnosed patients that on preliminary review would be appropriate candidates for the PGDP. These missing patients may outnumber known PGDP patients by a factor of 20 to 1 and, crucially, have a higher proportion of URM.

Overall, our findings demonstrate the presence of significant barriers to enrollment in the PGDP that work against all patients and families, but likely have a greater effect on URM as evidenced by their lower representation in referral routes requiring more steps. Focusing on relatively less biased routes of referral, such as from inpatient admissions, may improve the participation of URM in the PGDP. In this regard, actively using the EHR screening approach developed here to recruit additional patients is expected to help mitigate disparities; we are actively working to further refine this tool before implementation (see Dissemination). Finally, the results of this study are applicable and useful for other programs that grapple with similar challenges.

DISSEMINATION / ONGOING WORK

Plans for publication

From the data presented in this thesis, we plan to submit two manuscripts for peer-reviewed publication in Spring of 2023. The first will be a mixed-methods study on the barriers of undiagnosed families and the range of tools they use to navigate the odyssey. Questionnaire and interview data from Aim 1 will be supported with the PGDP cohort data in Aim 2, with a specific analysis for the experiences of URM families. Our second study will follow Aim 3, and will be an epidemiological analysis of the "missing" children with undiagnosed diseases.

Internal PGDP quality improvement

The findings from this thesis have been presented internally to the PGDP team, and changes are underway to improve access to the program, particularly for URMs. More specifically, PGDP is looking at ways to reduce the number of steps patients need to take and make the enrollment process less intimidating. One key quality improvement project, being directly driven by this medical student investigator, is the adoption of an electronic consent system.

Ongoing Work - Electronic consent system

One of the most significant barriers noted in by families is the lack of knowledge of PGDP's existence, how PGDP can help families, and what the requirements for referral are. We think that many of these issues, at least in part, are caused by the

current consent process. Currently, prospective participants are given the standardized traditional paper consent. We think this paper consent process is unnecessarily filled with technical jargon, time consuming, emotionally overwhelming and do not lead to a better understanding of what they are actually consenting to.[52] Therefore, we have worked on developing an easy-to-understand, video-based electronic consent system.

Electronic consent systems are a heterogenous group of new consent systems that typically involve a multimedia component.[52] Online video system increases knowledge retention, increase confidence, and decreases anxiety among parents.[53,54] Furthermore, they may also improve communication to those with audiovisual or cognitive disabilities.[55]

We scripted, storyboarded, and commissioned an animated video inspired by previous work for cerebral palsy genetic research enrollment at Phoenix Children's Hospital in Phoenix, Arizona. We decided on several key criteria for the design of the video:

- the characters must be visually diverse in race/ethnicity;
- the video needed to include children with visible disabilities;
- the format needs to be amenable to translation to other languages in the future.

Our video was submitted to Yale University's Institutional Review Board alongside other materials detailing the new online consent process for a preliminary review. A draft website was made by the Yale Information Technology Service (**Fig. 5**).

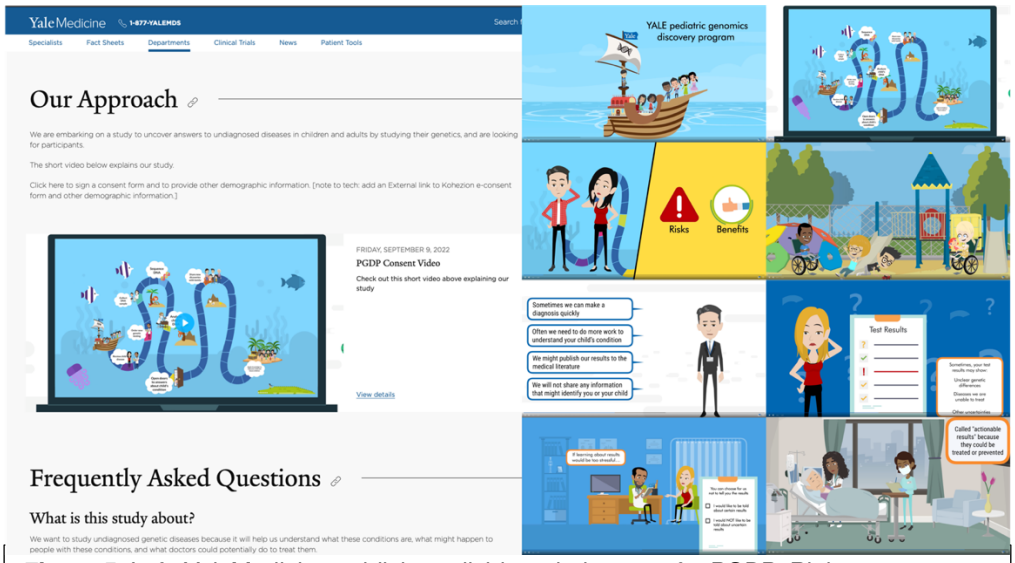

Figure 5. Left: YaleMedicine publicly available website page for PGDP. Right: Screenshots of the electronic consent video. A link to the video is available: https://yale.box.com/s/40qr07x85e81ukoe6efrxln15qzjc6zj.
A Spanish translation of the script has already been made, and we are in the process of making a Spanish-language dub.

We recognize there are pitfall to video-based consent. Potential participants may also want something that is portable and permanent so that they can revisit the information later on, and have interaction with the research team to address more individualized concerns.[55] Other barriers to access for e-consent systems include access to computers/internet, computer literacy, privacy concerns.[55] Our electronic consent system is designed around these concerns in the following ways:

- the website with the program information, consent form, and informational consent video will be publicly available at all times. As well, all participants will receive a copy of the full consent form in paper or electronic form, based on their preference;
- the electronic consent form will not replace a formal assessment by a PGDP staff member. The staff member will continue to review the consent and answer any questions during the initial evaluation;

- lastly, although the perception of risk of leaking private information is present, we use a HIPAA-compliant electronic consent system that complies with all federal and institutional regulations for identifiable health information handling. Participants will also be able to choose to defer registration until a formal meeting with a PGDP staff member.

Social media outreach and other organizations

Our participant interviews also noted that a contributor to the lack of knowledge about PGDP is a lack of brand awareness. For that reason, we are planning to increase visibility on the internet and social media platforms. Specific plan involves search engine optimization for web content, collaboration with rare/undiagnosed disease patient social media groups, and form close partnerships with PGDP families that are enthusiastic to serve as ambassadors for the program. Further, we were advised to find potential families in spaces that undiagnosed families frequent. Many of the families in our interviews sought support from schools, social services from the state and federal government, and had close relationships with genetic specialists. Thus, the next arm of our recruitment plan will be to reach out to these mediators who see a high proportion of undiagnosed families.

REFERENCES

1. Rare Diseases Act of 2002. Public Law 107-280. 2 ed: Office of the Federal Registrar, National Archives and Records Administration; 2002.
2. Moliner AM, Waligora J. The European Union Policy in the Field of Rare Diseases. Adv Exp Med Biol 2017;1031:561-587. DOI: 10.1007/978-3-319-67144-4_30.
3. Nguengang Wakap S, Lambert DM, Olry A, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet 2020;28(2):165-173. DOI: 10.1038/s41431-019-0508-0.
4. Genes G. Rare Disease Facts. Global Genes. (https://globalgenes.org/rare-disease-facts/).
5. Navarrete-Opazo AA, Singh M, Tisdale A, Cutillo CM, Garrison SR. Can you hear us now? The impact of health-care utilization by rare disease patients in the United States. Genet Med 2021;23(11):2194-2201. DOI: 10.1038/s41436-021-01241-7.
6. Chung CCY, Hong Kong Genome P, Chu ATW, Chung BHY. Rare disease emerging as a global public health priority. Front Public Health 2022;10:1028545. DOI: 10.3389/fpubh.2022.1028545.
7. Bogart K, Hemmesch A, Barnes E, et al. Healthcare access, satisfaction, and health-related quality of life among children and adults with rare diseases. Orphanet J Rare Dis 2022;17(1):196. DOI: 10.1186/s13023-022-02343-4.
8. Orphanet. Orphadata - Orphanet datasets. (https://www.orphadata.com).
9. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 2009;106(45):19096-101. DOI: 10.1073/pnas.0910672106.
10. Lakhani SA, Pierce R. Rethinking what constitutes a diagnosis in the genomics era: a critical illness perspective. Curr Opin Pediatr 2019;31(3):317-321. DOI: 10.1097/MOP.0000000000000754.
11. Meng L, Pammi M, Saronwala A, et al. Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management. JAMA Pediatr 2017;171(12):e173438. DOI: 10.1001/jamapediatrics.2017.3438.
12. Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. Science 2022;376(6588):44-53. DOI: 10.1126/science.abj6987.
13. Wood V, Lock A, Harris MA, Rutherford K, Bahler J, Oliver SG. Hidden in plain sight: what remains to be discovered in the eukaryotic proteome? Open Biol 2019;9(2):180241. DOI: 10.1098/rsob.180241.
14. Tangye SG, Al-Herz W, Bousfiha A, et al. Human Inborn Errors of Immunity: 2022 Update on the Classification from the International Union of Immunological Societies Expert Committee. J Clin Immunol 2022;42(7):1473-1507. DOI: 10.1007/s10875-022-01289-3.
15. Hussain SB, Quittner AL, Brown M, Li-Rosi AM. Understanding access to genomics in an ethnically diverse south Florida population: A comparison of

demographics in odyssey and rapid whole genome sequencing programs. J Genet Couns 2020;29(4):553-561. DOI: 10.1002/jgc4.1281.
16. Platt CD, Zaman F, Bainter W, et al. Efficacy and economics of targeted panel versus whole-exome sequencing in 878 patients with suspected primary immunodeficiency. J Allergy Clin Immunol 2021;147(2):723-726. DOI: 10.1016/j.jaci.2020.08.022.
17. Bauskis A, Strange C, Molster C, Fisher C. The diagnostic odyssey: insights from parents of children living with an undiagnosed condition. Orphanet J Rare Dis 2022;17(1):233. (In eng). DOI: 10.1186/s13023-022-02358-x.
18. Spillmann RC, McConkie-Rosell A, Pena L, et al. A window into living with an undiagnosed disease: illness narratives from the Undiagnosed Diseases Network. Orphanet J Rare Dis 2017;12(1):71. (In eng). DOI: 10.1186/s13023-017-0623-3.
19. Genes G. Annual Impact Report. 2020.
20. Gonzaludo N, Belmont JW, Gainullin VG, Taft RJ. Estimating the burden and economic impact of pediatric genetic disease. Genet Med 2019;21(8):1781-1789. DOI: 10.1038/s41436-018-0398-5.
21. Lavelle TA, Feng X, Keisler M, et al. Cost-effectiveness of exome and genome sequencing for children with rare and undiagnosed conditions. Genet Med 2022;24(6):1349-1361. DOI: 10.1016/j.gim.2022.03.005.
22. Incerti D, Xu XM, Chou JW, Gonzaludo N, Belmont JW, Schroeder BE. Cost-effectiveness of genome sequencing for diagnosing patients with undiagnosed rare genetic diseases. Genet Med 2022;24(1):109-118. DOI: 10.1016/j.gim.2021.08.015.
23. Schroeder BE, Gonzaludo N, Everson K, et al. The diagnostic trajectory of infants and children with clinical features of genetic disease. NPJ Genom Med 2021;6(1):98. DOI: 10.1038/s41525-021-00260-2.
24. McColley SA, Martiniano SL, Ren CL, et al. Disparities in first evaluation of infants with cystic fibrosis since implementation of newborn screening. J Cyst Fibros 2022 (In eng). DOI: 10.1016/j.jcf.2022.07.010.
25. Foundation CN. Shortening the Diagnostic Odyssey: 2020 Assessment Survey Summary. https://www.childneurologyfoundation.org: Child Neurology Foundation; 2020.
26. Walley NM, Pena LDM, Hooper SR, et al. Characteristics of undiagnosed diseases network applicants: implications for referring providers. BMC Health Serv Res 2018;18(1):652. DOI: 10.1186/s12913-018-3458-2.
27. Al-Ali S, Jeffries L, Faustino EVS, et al. A retrospective cohort analysis of the Yale pediatric genomics discovery program. Am J Med Genet A 2022;188(10):2869-2878. (In eng). DOI: 10.1002/ajmg.a.62918.
28. Hamosh A, Wohler E, Martin R, et al. The impact of GeneMatcher on international data sharing and collaboration. Hum Mutat 2022;43(6):668-673. DOI: 10.1002/humu.24350.
29. Splinter K, Adams DR, Bacino CA, et al. Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. N Engl J Med 2018;379(22):2131-2139. DOI: 10.1056/NEJMoa1714458.

30. Fraiman YS, Wojcik MH. The influence of social determinants of health on the genetic diagnostic odyssey: who remains undiagnosed, why, and to what effect? Pediatr Res 2021;89(2):295-300. DOI: 10.1038/s41390-020-01151-5.
31. Bonkowsky JL, Wilkes J, Bardsley T, Urbik VM, Stoddard G. Association of Diagnosis of Leukodystrophy With Race and Ethnicity Among Pediatric and Adolescent Patients. JAMA Netw Open 2018;1(7):e185031. DOI: 10.1001/jamanetworkopen.2018.5031.
32. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature 2016;538(7624):161-164. DOI: 10.1038/538161a.
33. von der Lippe C, Diesen PS, Feragen KB. Living with a rare disorder: a systematic review of the qualitative literature. Mol Genet Genomic Med 2017;5(6):758-773. (In eng). DOI: 10.1002/mgg3.315.
34. Deuitch NT, Beckman E, Halley MC, et al. "Doctors can read about it, they can know about it, but they've never lived with it": How parents use social media throughout the diagnostic odyssey. J Genet Couns 2021;30(6):1707-1718. DOI: 10.1002/jgc4.1438.
35. McConkie-Rosell A, Hooper SR, Pena LDM, et al. Psychosocial Profiles of Parents of Children with Undiagnosed Diseases: Managing Well or Just Managing? J Genet Couns 2018;27(4):935-946. DOI: 10.1007/s10897-017-0193-5.
36. Baumbusch J, Mayer S, Sloan-Yip I. Alone in a Crowd? Parents of Children with Rare Diseases' Experiences of Navigating the Healthcare System. J Genet Couns 2018 (In eng). DOI: 10.1007/s10897-018-0294-9.
37. Gengler AM. "I want you to save my kid!": Illness management strategies, access, and inequality at an elite university research hospital. J Health Soc Behav 2014;55(3):342-59. (In eng). DOI: 10.1177/0022146514544172.
38. Stanarević Katavić S. Health information behaviour of rare disease patients: seeking, finding and sharing health information. Health Info Libr J 2019;36(4):341-356. (In eng). DOI: 10.1111/hir.12261.
39. Palmer CGS, McConkie-Rosell A, Holm IA, et al. Understanding Adult Participant and Parent Empowerment Prior to Evaluation in the Undiagnosed Diseases Network. J Genet Couns 2018;27(5):1087-1101. DOI: 10.1007/s10897-018-0228-6.
40. Nambot S, Thevenon J, Kuentz P, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. Genet Med 2018;20(6):645-654. DOI: 10.1038/gim.2017.162.
41. Lee B, Hollenbeck-Pringle D, Goldman V, Biondi E, Alverson B. Are Caregivers Who Respond to the Child HCAHPS Survey Reflective of All Hospitalized Pediatric Patients? Hosp Pediatr 2019;9(3):162-169. DOI: 10.1542/hpeds.2018-0139.
42. Farooqi A, Jutlla K, Raghavan R, et al. Developing a toolkit for increasing the participation of black, Asian and minority ethnic communities in health and social care research. BMC Med Res Methodol 2022;22(1):17. DOI: 10.1186/s12874-021-01489-2.

43. Bureau USC. 5-Year American Community Survey, 2019. 2019.
44. Bureau USC. 5-Year American Community Survey Public Use Microdata Samples 2016-2020. 2020.
45. Aladangady A, Forde, A. Wealth Inequality and the Racial Wealth Gap. In: System BoGotFR, ed.2021.
46. Bureau USC. Median Family Income By Family Size. https://www.justice.gov/ust/eo/bapcpa/20220401/bci_data/median_income_table.htm2022.
47. Marin JR, Rodean J, Hall M, et al. Racial and Ethnic Differences in Emergency Department Diagnostic Imaging at US Children's Hospitals, 2016-2019. JAMA Netw Open 2021;4(1):e2033710. DOI: 10.1001/jamanetworkopen.2020.33710.
48. Aydin E, Torlak N, Haberman B, Lim FY, Peiro JL. The Survivorship Bias in Congenital Diaphragmatic Hernia. Children (Basel) 2022;9(2). DOI: 10.3390/children9020218.
49. Kumar P, Radhakrishnan J, Chowdhary MA, Giampietro PF. Prevalence and patterns of presentation of genetic disorders in a pediatric emergency department. Mayo Clin Proc 2001;76(8):777-83. (In eng). DOI: 10.1016/s0025-6196(11)63220-5.
50. Pelletier JH, Rakkar J, Au AK, Fuhrman D, Clark RSB, Horvat CM. Trends in US Pediatric Hospital Admissions in 2020 Compared With the Decade Before the COVID-19 Pandemic. JAMA Netw Open 2021;4(2):e2037227. DOI: 10.1001/jamanetworkopen.2020.37227.
51. Pique L, Graham S, Pearl M, Kharrazi M, Schrijver I. Cystic fibrosis newborn screening programs: implications of the CFTR variant spectrum in nonwhite patients. Genet Med 2017;19(1):36-44. DOI: 10.1038/gim.2016.48.
52. Nishimura A, Carey J, Erwin PJ, Tilburt JC, Murad MH, McCormick JB. Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized control trials. BMC Med Ethics 2013;14:28. DOI: 10.1186/1472-6939-14-28.
53. Book F, Goedeke J, Poplawski A, Muensterer OJ. Access to an online video enhances the consent process, increases knowledge, and decreases anxiety of caregivers with children scheduled for inguinal hernia repair: A randomized controlled study. J Pediatr Surg 2020;55(1):18-28. DOI: 10.1016/j.jpedsurg.2019.09.047.
54. Rothwell E, Wong B, Rose NC, et al. A randomized controlled trial of an electronic informed consent process. J Empir Res Hum Res Ethics 2014;9(5):1-7. DOI: 10.1177/1556264614552627.
55. Simon CM, Schartz HA, Rosenthal GE, Eisenstein EL, Klein DW. Perspectives on Electronic Informed Consent From Patients Underrepresented in Research in the United States: A Focus Group Study. J Empir Res Hum Res Ethics 2018;13(4):338-348. DOI: 10.1177/1556264618773883.
56. Luksic D, Sukhu R, Koval C, et al. A qualitative study of Latinx parents' experiences of clinical exome sequencing. J Genet Couns 2020;29(4):574-586. (In eng). DOI: 10.1002/jgc4.1276.

APPENDIX A: QUESTIONNAIRE OF BARRIERS ALONG THE DIAGNOSTIC ODYSSEY

[Consent]

I would like to potentially be contacted for an interview about my experiences (if selected, interview participants will be given an additional $50 Amazon gift card upon completion):

Yes | No

Questionnaire

Part 1: Demographics

Relation to patient	Parent | Sibling | Other family member | Unrelated | I am the patient
Race:	White | African American | Asian | Other
Hispanic:	Yes | No
Place of residence:	In CT | Outside CT
Gender Identification:	[Free text]
Marital Status:	Not married, living alone | Not married, cohabiting | Married | Divorced | Widowed
Education: (Highest Level Attained)	No formal schooling | Less than primary school | Primary school | Secondary school | College/University (Undergraduate degree) | College/University (Graduate degree)
Primary Language at Home:	English | Spanish | Other

Part 2: Barriers

What were the most important factors that delayed YOU from reaching our program? Please rank each item on a scale from 1 to 4:

1	2	3	4
Not a barrier or This did not cause any delays for me	A minor barrier or This may have delayed me by less than 7 days	A moderate barrier or This may have delayed me by less than 4 weeks	A major barrier or This may have delayed me by months to years

Access to healthcare in general

- ☐ I did not suspect the medical problems could be caused by a genetic condition
- ☐ I had a hard time scheduling appointments with my primary doctor or specialists
- ☐ My healthcare providers or I did not know this program existed
- ☐ I knew about the program, but my healthcare providers or I did not think our medical condition would qualify for this program, and/or wanted to wait for other diagnostic tests first
- ☐ I had trouble communicating with healthcare providers because we spoke different languages[56]

Mistrust or concerns about the program

- ☐ I knew about the program, but I had concerns about genetic testing
- ☐ I knew about the program, but I did not trust the program
- ☐ I knew about the program, but I was not sure how to join
- ☐ I knew about the program, but I was concerned with how much this program would cost
- ☐ I knew about the program, but I did not think the program would be helpful to us
- ☐ I knew about the program, but other family members did not want to participate

Logistics

- ☐ I had difficulty arranging for time off work to come here
- ☐ I had difficulty arranging for childcare to come here
- ☐ I had difficulty arranging for transportation to come here
- ☐ I was worried that participation would cause legal issues for me or my family

Other

- ☐ Other: (please specify)

Rank: These are the factors that you said caused Moderate or Major delays to reaching our program. Please rank **the top three factors** that caused the most delays from 1 to 3, **with 1 being the highest**.

APPENDIX B: SEMI-STRUCTURED INTERVIEW GUIDE

Interview Guide

Research question: What are the barriers to care to reaching PGDP for patients along the diagnostic odyssey?

Intent: The goal of section A is to explain the purpose of the semi-structured interview and to explain the role of the interviews in our larger agenda around this topic.

A. Introduction

a. Interviewer introduces themselves

b. Explain the purpose of the semi-structured interview and its role in the larger research agenda

c. Answer any participant questions

Intent: The intent of section B is to understand what factors may facilitate or impede patient or patient caregivers from reaching the program.

B. Reaching the program

- **What do you understand about the program, in your own words?**
- Can you describe how you learned about our program? (refer to questionnaire)
- Can you tell me about the barriers you had to coming here (refer to questionnaire)
    - How did you overcome these barriers?
- Why do you think some people would have a harder time reaching our program compared to others?
- What are some things that you think would turn people away from our program?
    - What do you think makes people want to participate in our program?
    - What is this going to do if this doesn't change how our child is going to be treated?
- Prompts if patients mention genetics:
    - Can you tell me about any experiences you may have had with genetic research? If you do not have any direct experiences, can you share things you've heard about it?

Intent: The intent of section C is to understand the interviewee's experiences with and perceptions of the general healthcare system, and how these interactions might affect participation in our program.

C. Access to healthcare

- Is there any kind of additional information or support that would have been helpful for you to have at any point in the referral to our program? (If yes) What kinds?[56]
- Can you tell me a little bit about your interactions with medical providers along the way, or those that referred you? What were things they did well or poorly?
    - Prompt: if they only mention negatives, ask about positives, and vice versa
- How does your experience with this program compare to other healthcare experiences you've had?

Intent: The intent of section D is to explore barriers to participation that might be unique to certain communities, such as people of colour.

D. Community

- What advice would you give yourself at the beginning of this? What do you wish you would have known about this process?
- Looking back, what are some things that could have helped other families in similar positions to yours?
- If people are not getting referral to us from genetics department referrals or the hospital, how do you think they could get here?
- If you are just starting the program, or just considering it, what are some things about our program/process that you would want to know?
- (General) We have some people who don't complete the process to enroll. Why do you think that is? What can we do about it?
    - (With minorities) We have a lot of trouble reaching [African American/Hispanic] patients who we think could benefit from this program. Why do you think that is? What can we do about it?
- Satisfied with the program?

Intent: The intent of section E is to close the interview by thanking the participant for their feedback and offering the opportunity to share any other thoughts they may have on topics that were not covered

E. **Session Closing**

a. Thank participants for their feedback and overall participation
b. Before we close the interview, are there things you think we didn't cover today that are important that we know about?

APPENDIX C: ICD-10-CM CODES USED IN INITIAL SCREEN FOR UNDIAGNOSED CHILDREN

D610; D6101; D6109; D613; D61818; D6182; D6189; D6182; D6189; D619; D640; D641; D643; D644; D6489; D690; D691; D692; D693; D694; D6941; D6942; D6949; D698; D699; D700; D704; D708; D709; D75; D761; D800; D801; D803; D804; D805; D806; D807; D808; D809; D829; D822; D849; D899; E25; E260; E271; E779; E799; E802; E880; E8801; E8802; E8809; E881; E882; E883; E884; E8840; E8841; E8842; E8849; E888; E8889; E889; F78; F842; F843; F848; F849; G11; G12; G241; G249; G250; G252; G253; G255; G2569; G258; G2581; G2582; G2583; G2589; G259; G3182; G32; G40; G60; G702; G71; G723; G729; G80; G900; G9009; G901; G903; G904; G908; G909; G910; G911; G912; G914; G918; G919; H40; H49; I420; I421; I422; I423; I424; I425; I428; I429; I4581; I6785; I780; J43; K72; K73; K753; K759; K76; K861; K8681; M04; M260; M269; M41; N07; O336; O337; P091; P092; P093; P095; P096; P098; P099; P941; P942; P948; P949; P960; Q00; Q01; Q03; Q04; Q06; Q07; Q11; Q12; Q13; Q14; Q15; Q20; Q212; Q213; Q214; Q218; Q219; Q22; Q23; Q24; Q251; Q252; Q2521; Q2529; Q254; Q2540; Q2541; Q2542; Q2543; Q2544; Q2545; Q2546; Q2547; Q2548; Q2549; Q333; Q334; Q335; Q336; Q338; Q339; Q349; Q402; Q403; Q408; Q409; Q41; Q42; Q432; Q433; Q446; Q447; Q450; Q451; Q453; Q458; Q459; Q500; Q5001; Q5002; Q503; Q5031; Q5032; Q5039; Q506; Q51; Q520; Q521; Q5210; Q5211; Q5212; Q52120; Q52121; Q52122; Q52123; Q52124; Q52129; Q524; Q526; Q527; Q5270; Q5271; Q5279; Q528; Q529; Q550; Q551; Q5520; Q5521; Q5523; Q5529; Q553; Q554; Q555; Q557; Q558; Q559; Q56; Q600; Q601; Q602; Q603; Q604; Q605; Q611; Q6111; Q6119; Q612; Q613; Q614; Q615; Q618; Q619; Q621; Q622; Q623; Q624; Q625; Q626; Q627; Q628; Q63; Q645; Q71; Q72; Q73; Q74; Q750; Q751; Q752; Q754; Q755; Q758; Q759; Q760; Q761; Q762; Q763; Q764; Q7641; Q76411; Q76412; Q76413; Q76414; Q76415; Q76419; Q7642; Q76425; Q76426; Q76427; Q76428; Q76429; Q7649; Q766; Q767; Q768; Q769; Q77; Q78; Q808; Q809; Q81; Q858; Q859; Q870; Q871; Q8711; Q8719; Q872; Q873; Q875; Q878; Q8781; Q8782; Q8789; Q89; R26; R4183; R41842; R41843; R569; R651; R7983; Z94; E22; E23; E283; E70; E71; E72; E74; E8881; F840; F845; G23; J44; J84; N05; N18; N26; N27; P91; Q05; Q16; Q17; Q26; Q27; Q28; Q300; Q301; Q302; Q308; Q309; Q31; Q32; Q35; Q36; Q37; Q39; Q67; Q69; Q70; Q83; R27; R6252; Z15

APPENDIX D: THE FIRST 10 UNDIAGNOSED PATIENTS FOUND BY OUR SCREENING METHOD

Diagnosis	Notes	Age	Sex	Hispanic or Latino	Race
Autistic disorder (F840)	Autism, long QT, Extended gene analysis found suspected 3-exon deletion in TRDN gene but no explanation for autism	9	Male	Hispanic or Latino	Other/Not Listed
Unsp lack of expected normal physiol dev in childhood (R6250);	Congenital nystagmus, clonic seizure, Cerebral palsy, global developmental delay, G-tube dependence. Had muscle biopsy done to look for mitochondrial disorder, nothing was found. Followed by spencer-manzon	5	Female	Non-Hispanic	Black or African American
Other specified congenital malformations of spinal cord (Q068);	Congenital omphalocele, cloacal anomoly, lipomeningocele with tethered cord, congenital hip dysplasia	3	Male	Non-Hispanic	White or Caucasian
Congenital hypotonia (P942);	Congenital hypotonia, dysmorphic features, moderate atopy, negative extended gene analysis	1	Female	Non-Hispanic	Asian
Spastic quadriplegic cerebral palsy (G800);	Failure to thrive with height/weight 3rd percentile, cryptorchidism	3	Male	Non-Hispanic	Other/Not Listed
Congenital malformations of corpus callosum (Q040);Unsp lack of expected normal physiol dev in childhood (R6250);	Corpus collosum malformation with seizures and developmental delay. Has a de novo ZBTB18 gene mutation	0	Female	Non-Hispanic	White or Caucasian
Epileptic spasms, not intractable, w/o status epilepticus (G40822);Congenital hypotonia (P942);Other reduction deformities of brain (Q043);Unsp lack of expected normal physiol dev in childhood (R6250);	Miller dieker syndrome [Clinical or genetic diagnosis?]	1	Female	Non-Hispanic	White or Caucasian
Autistic disorder (F840);Unsp lack of expected normal physiol dev in childhood (R6250);	Autism, BMI >99%ile, scoliosis, congenital talipes equinovarus, amelogenesis imperfecta (enamel defect)	0	Male	Hispanic or Latino	Other/Not Listed
Spastic quadriplegic cerebral palsy (G800);Cerebral palsy, unspecified (G809);	Global delay, ex26wk, GT dependent, asthma, recurrent infections requiring PICU, PVCs, sensorineural hearing loss, blindness	2	Male	Hispanic or Latino	White or Caucasian

Delayed milestone in childhood (R620);	Global delay, ADHD. 22q12.3 deletion does not sufficiently explain global delay and behavioral issues. Expressed interest in WES	1	Male	Non-Hispanic	White or Caucasian