American Journal of Medical Genetics Part A

WILEY

AMERICAN JOURNAL OF medical genetics PART A

ORIGINAL ARTICLE

Generating Advancements in Longitudinal Analysis in X and Y Variations: Rationale, Methods, and Diagnostic Characteristics for the GALAXY Registry

Alexandra Carl[1,2] | Samantha Bothwell[2] | Karli Swenson[2] | Ryan Bregante[1,3] | Lilian Cohen[1,4] | Virginia Cover[1,5] | Anna Dawczyk[1,6] | Gail Decker[1,7] | Stephen B. Gerken[1] | David Hong[1,8] | Susan Howell[1,2,9] | Armin Raznahan[1,10] | Alan D. Rogol[1,11] | Nicole Tartaglia[1,2,9] | Shanlee Davis[1,2,12]

[1]Generating Advancements in Longitudinal Analysis in X and Y Variations (GALAXY) Consortium | [2]eXtraOrdinarY Kids Clinic and Research Program, Children's Hospital Colorado, Aurora, Colorado, USA | [3]Living With XXY, San Diego, California, USA | [4]Department of Pediatrics, Division of Medical Genetics, Weill Cornell Medical College, New York, New York, USA | [5]The Association for X and Y Chromosome Variations, Church Hill, Maryland, USA | [6]Department of Sociology, The University of British Columbia, Vancouver, British Columbia, Canada | [7]The XXYY Project, Church Hill, Maryland, USA | [8]Department of Psychiatry and Interdisciplinary Brain Sciences, Stanford University, Stanford, California, USA | [9]Division of Developmental Pediatrics, Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado, USA | [10]Developmental Neurogenomics Unit, Child Psychiatry Branch, National Institute of Mental Health, Bethesda, Maryland, USA | [11]Pediatric Diabetes and Endocrinology, University of Virginia, Charlottesville, Virginia, USA | [12]Division of Diabetes and Endocrinology, Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado, USA

Correspondence: Alexandra Carl (alexandra.carl@cuanschutz.edu)

Received: 3 December 2024 | **Revised:** 23 July 2025 | **Accepted:** 26 July 2025

Funding: This work was supported by Association for X and Y Chromosome Variations, Living with XXY, The XXYY Project, and National Center for Advancing Translational Sciences, UM1 TR004399.

Keywords: Klinefelter syndrome | registry | sex chromosome aneuploidies | trisomy X | XXY | XXYY | XYY

ABSTRACT

Sex chromosome aneuploidies (SCAs) are a family of genetic disorders that result from an atypical number of X and/or Y chromosomes. SCAs are the most common chromosomal abnormality, affecting ~1/400 live births, yet are often underdiagnosed, leading to over-representation of more severely impacted individuals in many clinical studies. In addition to this ascertainment bias, existing work in SCAs has also been limited by low geographic and demographic diversity. To address these limitations, we have created the Generating Advancements with Longitudinal Analysis in X and Y variations (GALAXY) Registry. Through prioritizing sustainability, transparency, and minimizing participant burden, the overarching goal of the GALAXY Registry is to improve health outcomes for individuals with SCAs by serving as an infrastructure for future SCA research based on a large, heterogeneous, and longitudinal sample. To date, GALAXY has accrued 335 verified SCA participants with an average accrual of 11.2 participants/month (6.7 47,XXY, 1.9 47,XXX, 2.0 47,XYY, 3.2 48,XXYY, 1.8 48,XXXY, and 1.3 Other). Demographic data between those identified to have SCA prenatally (predominantly cell-free DNA screening) differ from those diagnosed postnatally for insurance status, age at enrollment, genetic test type, and reason for SCA diagnosis. Next steps include targeted recruitment of underrepresented groups (e.g., non-47,XXY karyotypes, older adults, minoritized individuals), extraction of medical record data into the registry, international expansion, and continued engagement with the SCA community. As a collaboration between clinician investigators and the SCA community, the GALAXY Registry is a powerful resource for future patient-centered clinical research.

1 | Introduction

Sex chromosome aneuploidy (SCA) conditions refer to a family of genetic disorders resulting from an atypical number of X or Y chromosomes. Together, SCAs occur in 1 in ~400 live births, making them the most commonly occurring chromosomal abnormalities despite being under-recognized and underdiagnosed (Nielsen and Wohlert 1990). Supernumerary SCAs include Klinefelter syndrome (47,XXY; 1/600 males), Trisomy X syndrome (47,XXX; 1/1000 females), XYY syndrome (47,XYY; 1/1000 males), and rarer tetrasomy and pentasomy conditions (e.g., 48,XXYY, 48,XXXY, 48,XXXX, 49,XXXXY, etc. estimated to occur between 1 in 17,000 to 100,000 live births) (Linden et al. 1995; Nielsen and Wohlert 1990). Historically, SCAs have been under-ascertained, with just 10%–40% of individuals ever receiving a diagnosis, of whom only 10% are diagnosed in childhood (Abramsky and Chapple 1997; Sánchez et al. 2023; Tartaglia et al. 2020). Additionally, the diagnostic odyssey often starts with an observed medical problem leading to genetic testing, which introduces ascertainment bias such that individuals with more severe presentations are more likely to be diagnosed. However, the rate of prenatal identification is increasing as prenatal cell-free DNA (cfDNA) becomes part of routine prenatal care (Abramsky and Chapple 1997). With this increase in both earlier diagnosis and overall ascertainment of SCAs, there is an unprecedented demand for patient-centered research (Velvin et al. 2022) targeted at improving health and wellbeing for individuals affected by these conditions. Additionally, this increased prenatal diagnosis has decreased ascertainment bias in SCA diagnosis over the past few years and raises the question of how relevant existent literature is for those being incidentally identified today.

Several challenges hinder clinical research in SCAs. Foremost, SCAs are each rare, leading to an overrepresentation of underpowered studies on small cohorts at single institutions. Due to small numbers, SCA conditions are often pooled together for more analytical power, thus minimizing unique differences driven by karyotype or other genetic and environmental covariates that contribute to the phenotypic heterogeneity. While larger-scale retrospective studies have been possible using European registries (Berglund et al. 2022; Jørgensen et al. 2019), these are predominately studies of adults and lack racial, ethnic, and/or socioeconomic diversity that truly represents all persons living with SCA. Newborn screening studies from the 1960s–1970s prospectively followed 304 individuals with SCAs diagnosed incidentally at birth across cohorts in Colorado (Robinson et al. 1979), Denmark (Neilson et al. 1979), Toronto (Stewart et al. 1979), New Haven (Leonard et al. 1979), Tokyo (Higurashi et al. 1979), and Edinburgh (Ratcliffe et al. 1979). Although these studies provide a foundation for the natural history of growth and development in these conditions, they do not reflect current clinical practice and healthcare advancements over the past half-century.

To address some of these limitations, a collaborative effort between clinician investigators and community partners developed the Generating Advancements in Longitudinal Analysis in X and Y Variations (GALAXY) Registry that launched in 2022. This clinical registry uses a robust infrastructure based on a successful model in Turner syndrome—Inspiring New Science in Guiding Healthcare in Turner Syndrome (INSIGHTS) Registry (Shankar et al. 2024). The GALAXY registry is guided by a diverse Steering Committee (SC) of researchers, clinicians, self-advocates, and parents of affected individuals to inform the development, oversight, and direction of all GALAXY projects. Through prioritizing sustainability, transparency, and minimizing participant burden, the overarching goal of the GALAXY Registry is to improve health outcomes for individuals with SCAs by serving as an infrastructure for future SCA research based on a large, heterogeneous, and longitudinal sample. The purpose of this paper is to detail the development and methods of the GALAXY Registry, present data on the cohort to date, and explore differences between pre- and postnatal SCA ascertainment.

2 | Methods

2.1 | Structure of GALAXY

An overview of the GALAXY infrastructure is depicted in Figure 1. A Steering Committee (SC) composed of 13 collaborators—including clinicians, researchers, and family and self-advocates—was established to direct the GALAXY Registry and ensure that resource and data use reflect the interests and priorities of the SCA community. The initial GALAXY Steering Committee members were appointed by the Principal Investigator (PI) and leadership from the Association for X and Y Variations (AXYS). As additional members are needed due to attrition or recognition of a gap in expertise, they can be recommended by any committee member and voted in by a majority vote of the committee members. The role of the SC is to provide stakeholder input, consensus, and oversight for research projects conducted by approved study teams. The responsibilities of the SC include review, modification, and ultimate approval of study-related documents including, but not limited to, the study protocol, data collection instruments, and participant-facing materials (i.e., recruitment materials). SC members are all volunteers who serve a two-year term, meet monthly to establish goals and priorities, review new proposals, and oversee project progress. Members follow standard conflict of interest (COI) guidelines and are cleared of COI prior to approval of membership to the SC. In addition, smaller working groups are developed as needed to accomplish specific tasks, including a Diversity, Equity, and Inclusion (DEI) group, a Genetic/Medical group, and a Psych-Behavioral Health group.

The University of Colorado (CU) serves as the lead site and data coordinating center for the GALAXY Registry. CU is responsible for administration tasks related to the registry, including regulatory compliance through a Single Institutional Review Board (sIRB), development and maintenance of the database and Standard Operating Procedures, coordinating virtual SC meetings, onboarding new participant enrollment sites, and training of all study personnel. A timeline of study start-up can be seen in Figure 2. REDCap (Research Electronic Data Capture) (Harris et al. 2009) is used for both clinically verified data capture and participant-facing surveys in GALAXY. REDCap is a secure, HIPAA-compliant, web-based application designed to support data capture for research studies,

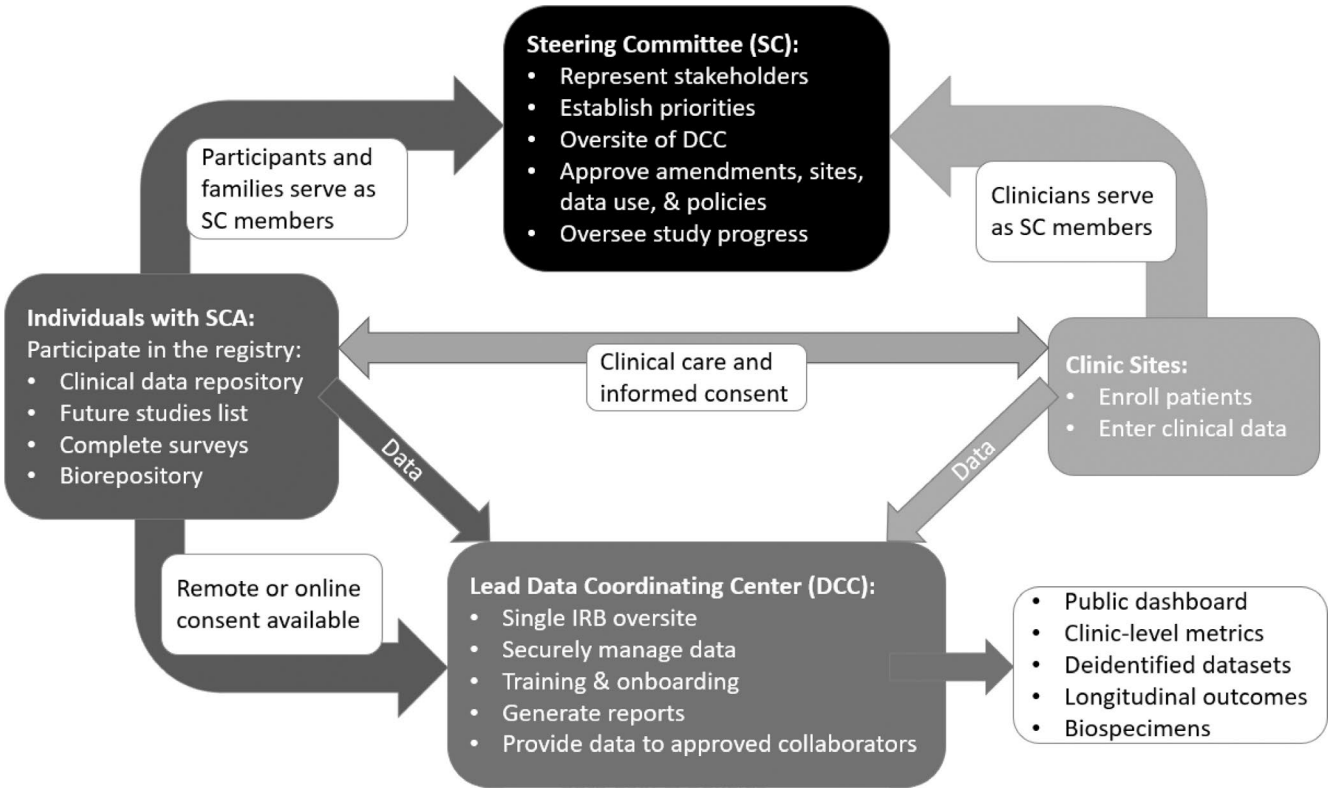

FIGURE 1 | Stakeholders involved in the GALAXY registry. Abbreviations: DCC = data coordinating center; IRB = Institutional Review Board; SC = Steering Committee; SCA = sex chromosome aneuploidy.

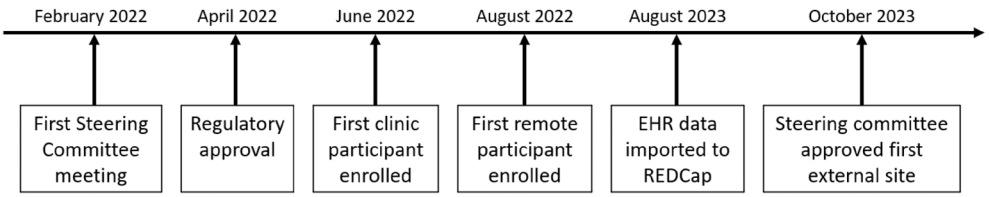

FIGURE 2 | Timeline of study-start up for GALAXY. EHR = electronic health record.

with a built-in e-consent framework and dual language management system to allow participants to consent in their preferred language (English or Spanish). The GALAXY Registry is approved by the Colorado Multiple Institutional Review Board (COMIRB, #20–0482).

Individuals with SCAs can enroll in GALAXY through a participating clinic site or online with appropriate genetic test results confirming an SCA. Participating sites are approved, multidisciplinary clinics that provide care to individuals with SCAs, agree to enroll participants in GALAXY, and have resources to contribute clinical data to the database. Applications require approval by the Steering Committee to ensure the site has the capacity to meaningfully contribute to the registry goals. Minimal requirements for participating sites include that the site must provide clinic care to patients with SCAs, have capacity for human subject research (such as appropriately trained personnel), and the ability to enroll and contribute data for at least 15 participants annually. The onboarding process includes the local IRB ceding to COMIRB, finalizing the Registry Agreement that authorizes data and material transfer between institutions, and training new site personnel.

2.2 | Recruitment and Enrollment

At participating clinics, potentially eligible patients are identified by clinic providers via review of past and upcoming clinic schedules by research staff, and/or via electronic health record (EHR) tools.

Inclusion criteria include:

1. Cytogenetically confirmed diagnosis of SCA (including mosaic karyotypes);
2. Any age, gender, or language; and
3. Providing informed consent for individuals ≥ 18 years of age, or parent/guardian consent for individuals < 18 or ≥ 18 with impaired decision making (plus assent for minors).

	In Clinic Process	*Remote Process*
Recruit	Identify eligible clinic patients	Advocacy organizations: newsletters, websites, social media, etc.
	Message sent via patient portal	Community events
	Discusses study with patient in clinic	Non-affiliated clinicians
Consent	Face-to-face informed consent and assent as applicable	Individual or caregiver review online consent form (video option)
		Comprehension quiz to ensure understanding
	Signature(s)	E-signature(s)
	Copy to participant	Downloadable PDF
Confirm Eligibility	EHR reviewed to confirm available genetic test results confirm SCA	Participant uploads a genetic test OR signs a Release of Records (ROR) form
		Obtain test results confirming SCA
Enroll	Collection of basic info from participant to create GUID	
	Collect blood, if consented	
Data Acquisition	Extraction of discrete elements from EHR	
	Manual data entry	
	ROR as needed	

FIGURE 3 | Recruitment and data acquisition processes. EHR = electronic health record; GUID = global unique identifier; ROR = release of records.

The only exclusion criterion is the lack of documentation of genetic testing confirming an SCA diagnosis or deceased individuals.

Enrollment processes are outlined in Figure 3. Eligible patients and/or their legal guardians are recruited by a member of their clinic site or research team, and informed consent and assent documents are available in both English and Spanish, with the opportunity to translate into other languages as a need arises. The SC intends that participating staff approach > 80% of eligible clinic patients about the study opportunity and obtain > 50% participation agreement.

Informed consent is obtained from participants 18 years and older who can consent for themselves. Informed consent is obtained from parents or legally authorized representatives of individuals under 18 years of age or those 18 years and older who are unable to consent for themselves due to developmental disability (as determined by the patient's clinician and/or previous legal determinations in accordance with the applicable law). Additionally, assent is obtained from those between 7 and 17 years and those 18 years and older who are unable to consent for themselves.

Eligible individuals also have the option to self-enroll if they are not seen in one of the participating sites. Self-enrollment allows for participation from individuals who do not have access to multidisciplinary SCA clinics or SCA research centers, reducing ascertainment bias of the registry. The GALAXY Registry is advertised on institution and organization websites and social media pages. Partnerships with SCA advocacy groups such as the Association of X and Y Variations (AXYS), The XXYY Project, and Living with XXY allow broader recruitment. Research staff attend and present the GALAXY registry during virtual community meetings and webinars to talk about the registry and inform potential participants of the study procedures. Additionally, SCA advocacy groups post online about the registry on their social media platforms (e.g., Facebook groups, Instagram) and on email newsletters. In July 2023, video consenting was approved for the registry due to the observed increased incidence of language-based difficulties in SCA populations (Lee et al. 2011). Videos are available alongside written online consent that breaks down and explains each section of the consent for individuals for whom reading may be a barrier.

To confirm eligibility, individuals enrolling online, and not seen within a participating clinic, are required to upload their genetic test results or sign a Release of Records for the research team to request a valid copy of their cytogenetic lab report. Participants (or their legal guardian) can securely upload additional applicable medical records or provide permission for records to be requested from clinics and released to the study team so that clinically verified data can be included in the registry database. Expanding recruitment to individuals who are not patients at a participating clinic site widens participation opportunities and has the potential to recruit a more diverse sample to the registry. In addition to allowing medical records to be used for research purposes, participants can consent to optional procedures, including being recontacted for future SCA-related research and receiving electronic research surveys. Individuals seen at a participating clinic can additionally consent to contributing a blood sample to the biorepository.

2.3 | Data Acquisition, Management, and Sharing

At the time of enrollment, participants provide demographic information, including race/ethnicity, insurance status, birth date, location of birth, and current residence. Birth, medical, and surgical history, diagnostic information, and medications are abstracted directly from medical records upon enrollment and updated annually. For participants who receive care from a participating institution, discrete data elements are automatically extracted monthly from the EHR into the research database. This EHR extraction script was developed and designed by the lead site to allow standardized implementation at other participating clinic sites. Data

that cannot be automatically exported are manually reviewed and entered into REDCap by the local study team. For health records outside the participating sites, a signed Release of Records is obtained during the consent process, and relevant medical records are requested, abstracted, and entered into the database manually. When data cannot be obtained from medical records, the study team contacts the participant for additional information. Data quality is ensured by trained research staff following Standard Operating Procedures, and data are validated during site onboarding and training. Participants consent to long-term follow-up through the registry, and data are updated yearly with no defined endpoint to participation.

Data sharing from the GALAXY Registry is managed following the Data Sharing Agreements (a legal agreement between institutions) and Data Use Policy (publicly available on galaxyregistry.org, governs the process for handling data use requests) approved by the Steering Committee. In addition to requesting data, investigators can apply to disseminate recruitment materials to participants who consented to the optional Future Studies Recruitment List (Figure 4). In both scenarios, the investigator would submit a Data Use Proposal (template publicly available on the galaxyregistry.org website). The Steering Committee reviews all proposals to ensure they are of sound scientific merit and beneficial to individuals and families in the SCA community. Data proposals are requested to include community engagement and describe planned dissemination to the SCA community, separate from academic conferences and peer-reviewed publication.

In addition, each participant is given a Global Unique Identifier (GUID) through the National Institutes of Health (NIH) Biomedical Research Informatics Computing System. The Rare Diseases Registry Program (RaDaR) provides guidance for setting up and maintaining registries for rare diseases. The GUID and common data elements taken from RaDaR facilitate data sharing without sharing personally identifying information (PII) across studies.

Sharing data outside the academic community is a goal of the GALAXY registry. A live dashboard with basic data is publicly available (on the galaxyregistry.org website) to increase transparency with the registry participants, prospective researchers, and the SCA community. Additionally, presentations and publications about the registry are shared on the study website (galaxyregistry.org). The consent form is also publicly available (redcap.link/GALAXY-consent) for individuals and prospective researchers to review.

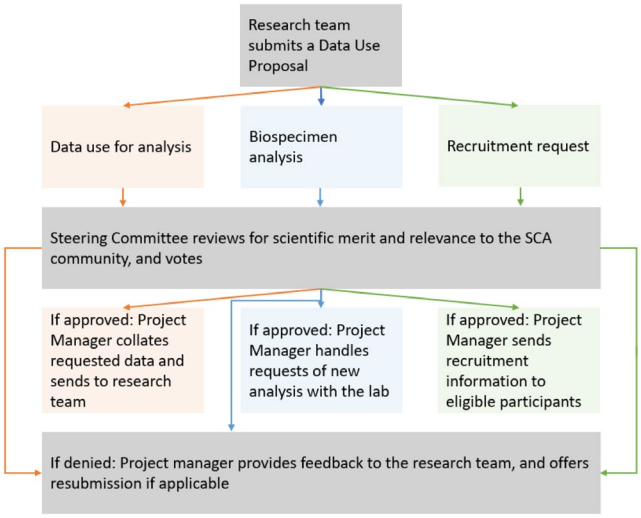

FIGURE 4 | How researchers can interact with GALAXY.

2.4 | Outcome Measures

Common data elements (CDEs) were selected from NIH repositories. Expanded and specialty-specific data elements were developed and refined by interdisciplinary members of the SC to ensure reliability and validity of the measures. Validated patient-facing measures are included (Supporting Information) and any non-validated questionnaires were reviewed and modified by the SC and pilot tested with individuals with SCAs.

Data elements in GALAXY, abstracted from participant's EHR:

1. Genetic test results.
2. Timing and reason for SCA diagnosis.
3. Other medical diagnoses (i.e., presence, number, and type).
4. Growth parameters and vital signs.
5. Documented physical exam including pubertal staging, dysmorphic features, and neurological exam.
6. Antenatal, birth and neonatal history.
7. Surgical procedures (i.e., presence, number, and type).
8. Hospitalizations (i.e., presence, number, reason).
9. Treatments including medications (i.e., presence, number, and type).
10. Imaging tests performed and results.
11. Laboratory tests performed and results.
12. Psychological and cognitive tests performed and results.

Additionally, from those who consent to participate in the optional questionnaires and biorepository:

13. Patient-reported outcomes from questionnaires in domains of quality of life, wellbeing, lived experiences, sociodemographic information, access to care, psychosocial health, and so forth. (Table S1).
14. Blood draw with plasma, serum, and buffy coat stored in the biorepository.

GALAXY surveys and information regarding future research opportunities are sent via REDCap to the email address the participant provided at enrollment. For participants under 18 years—or at the request of an adult participant—the emails are sent to a parent/guardian. Surveys are electronically administered utilizing the survey queue tool in REDCap, and questionnaires specific to the participant's age or karyotype are populated into their unique survey queue, which participants can return to at any time. Surveys are designed to collect participant-reported

outcomes including sociodemographic information, access and experience with clinical care, quality of life, wellbeing, and psychosocial health. Surveys include a statement indicating results are not shared with the participant's clinical providers and that participants should contact their doctor(s) directly if they have clinical concerns that arise from completing the survey.

2.5 | Statistical Analysis

Demographic and diagnostic summaries are presented as count (%). Age is summarized as median [IQR] due to non-normality, as tested with Shapiro-Wilks. Demographic and diagnostic history data are stratified by timing of SCA diagnosis (pre- vs. postnatal). Differences between pre- and postnatal diagnosis groups were tested using Chi-Squared tests, or Fisher's Exact tests if expected cell counts were less than 5. The difference in age between groups was tested using a Wilcoxon Rank Sum test. A type 1 error rate of 0.05 is assumed. All analyses were performed in R, version 4.4.0.

3 | Results

Children's Hospital Colorado (CHCO) began recruiting clinical patients in June of 2022; four more sites are currently

TABLE 1 | Sample demographic characteristics.

	Total ($n=301$)	Prenatal ($n=139$)	Postnatal ($n=162$)	p
Age at enrollment (years)	10.0 [1.9, 17.3]	2.0 [0.32, 9.3]	14.7 [9.8, 22.8]	**<0.001**
Enrollment group				**0.015**
CHCO	190 (63.1%)	98 (70.5%)	92 (56.8%)	
Online	111 (36.9%)	41 (29.5%)	70 (43.2%)	
Race and ethnicity				0.052
White non-Hispanic/Latinx	219 (72.8%)	94 (67.6%)	125 (77.2%)	
Hispanic/Latinx	41 (13.6%)	18 (12.9%)	23 (14.2%)	
Other non-Hispanic/Latinx	18 (6.0%)	10 (7.2%)	8 (4.9%)	
More than one race	12 (4.0%)	10 (7.2%)	2 (1.2%)	
Unknown	11 (3.7%)	7 (5.0%)	4 (2.5%)	
Insurance at enrollment				**<0.001**
Medicaid/Medicare	57 (18.9%)	16 (11.5%)	41 (25.3%)	
Private	180 (59.8%)	101 (72.7%)	79 (48.8%)	
Military	10 (3.3%)	7 (5.0%)	3 (1.9%)	
2 or more	23 (7.6%)	6 (4.3%)	17 (10.5%)	
No insurance	2 (0.7%)	0 (0%)	2 (1.2%)	
Other	9 (3.0%)	3 (2.2%)	6 (3.7%)	
Unknown	20 (6.6%)	6 (4.3%)	14 (8.6%)	
Optional procedures				
Future studies list	287 (95.3%)	130 (93.5%)	157 (96.9%)	0.181
Questionnaires	277 (92.0%)	126 (90.6%)	151 (93.2%)	0.516
Biobank	113 (37.5%)	60 (43.2%)	53 (32.7%)	0.065
Recruitment method				**0.006**
AXYS	33 (11.0%)	12 (8.6%)	21 (13.0%)	
Clinic staff or doctor	170 (56.5%)	92 (66.2%)	78 (48.1%)	
Community member	4 (1.3%)	2 (1.4%)	2 (1.2%)	
Living with XXY	59 (19.6%)	21 (15.1%)	38 (23.5%)	
Other	16 (5.3%)	2 (1.4%)	14 (8.6%)	
Social media	19 (6.3%)	10 (7.2%)	9 (5.6%)	

Note: Significance level = 0.05. Age differences were tested using a Wilcoxon Rank Sum test. Categorical differences were tested using Chi-Squared tests or Fisher's Exact tests if expected cell counts were less than 5. Age is summarized as median [interquartile range].

onboarding as of November 2024. A total of 325 individuals have been enrolled and validated based on enrollment criteria into the registry as of November 4th, 2024. Of the 325 enrolled, 301 had complete diagnostic information, 190 of which were from CHCO and 111 from self-enrollment with an average of 11 new participants per month. The median age at enrollment is 10 years (IQR: 1.9–17.3 years), ranging from 0 to 87 years. Demographic data can be seen in Table 1. As expected, participants who were diagnosed prenatally were significantly younger when they enrolled into GALAXY than participants diagnosed postnatally (median 2 [IQR: 0.32, 9.3] vs. 14.7 [IQR: 9.8, 22.8] years; $p < 0.001$). A higher proportion of prenatally diagnosed participants were enrolled at CHCO compared to participants who were postnatally diagnosed (70.5 vs. 56.8%; $p = 0.015$) and the distribution of recruitment method differed between groups ($p = 0.006$). A higher proportion of participants diagnosed prenatally were recruited by clinic staff or doctors (66.2% vs. 48.1%), which aligns with this group being more often enrolled at CHCO. A higher proportion of participants diagnosed postnatally were recruited through the non-profit advocacy partner Living with XXY (23.5% vs. 15.1%). However, there was not evidence of a difference in rates of consent to optional procedures (future studies list, questionnaires, biobank) between participants diagnosed prenatally vs. postnatally. There was sufficient evidence that the distribution of insurance at enrollment differed between groups ($p < 0.001$). A higher proportion of participants prenatally diagnosed were covered via private insurance (72.7% vs. 48.8%) whereas a higher proportion of participants postnatally diagnosed were covered by Medicaid/Medicare (25.3% vs. 11.5%, $p < 0.001$).

Forty-one total states are represented in the cohort, as visualized in Figure 5, as the state of residence for US-based participants. Maximum recruitment for a single state is 116 participants in Colorado. The average enrollment per month is 11.2 participants, with the greatest enrollment occurring in July 2023, which coincides with the Association for X and Y Variations Annual Conference (Figure 6). Average accrual per month for each SCA subgroup is 6.7 for 47,XXY, 1.9 47,XXX, 2.0 47,XYY, 3.2 48,XXYY, 1.8 48,XXXY, and 1.3 Other.

Diagnostic data can be seen in Table 2. Of the 301 participants with diagnostic information available, a higher proportion were diagnosed with their SCA postnatally (162, 53.8%) than prenatally (139, 46.2%). The distribution of SCA categories differed between the pre- and postnatally diagnosed groups ($p < 0.001$) such that a higher proportion of participants who were diagnosed prenatally have XXY (68.3% vs. 53.7%) while a higher proportion of participants who were diagnosed postnatally have XXYY (25.9% vs. 7.9%). Of those diagnosed postnatally, the median age at diagnosis is 7.2 years (IQR: 2.3–15.1 years), ranging from birth to 55 years. Genetic test type used for diagnosis was different between the two groups, with more prenatal diagnoses occurring via karyotype and FISH and more postnatal diagnoses occurring via microarray.

Figures 7 and 8 show reasons for genetic testing that resulted in the SCA diagnosis for prenatal and postnatal diagnoses, respectively. The most common reason for prenatal genetic testing was advanced maternal age (55, 40%) and the most common reasons for postnatal genetic testing were developmental delay (58, 36%), learning or behavioral concerns (31, 19%), and dysmorphic physical features (19, 12%). When the reasons for genetic testing are collapsed into incidental or secondary to symptoms, there was a difference between the groups, such as prenatal diagnoses being more often incidental (69.1%) while postnatal diagnoses were more often secondary to symptoms (96.3%; $p < 0.001$).

4 | Discussion

The GALAXY Registry is a collaborative effort among invested stakeholders to collect and store longitudinal clinical and self-reported data from individuals with SCA in a centralized international database with common, expanded, and specialized data elements. Since launching in April 2022, GALAXY has enrolled > 300 individuals from diverse genetic, geographical, racial, and socioeconomic populations and continues to enroll participants and collect data, with the expectation of serving as a sustainable resource for future natural history and other observational SCA studies. The diversity of timing and reason for diagnosis allows

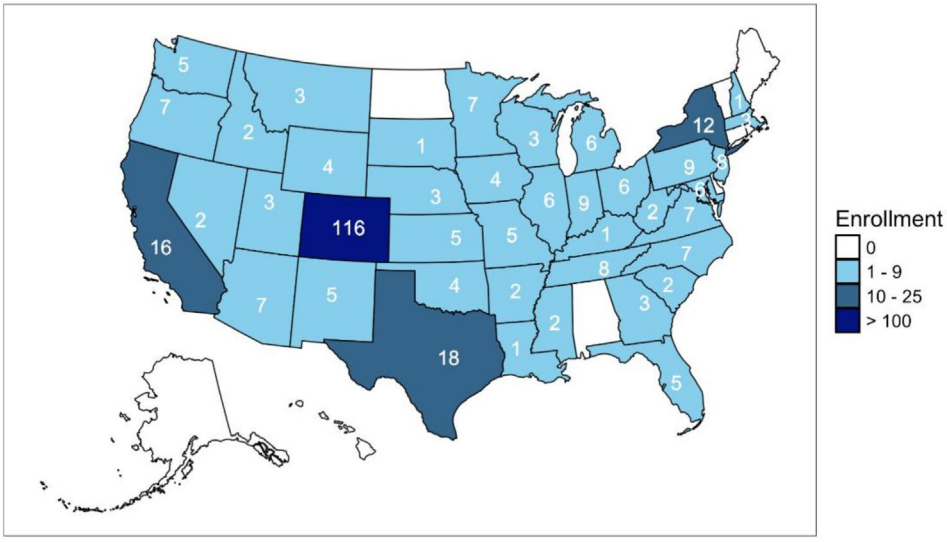

FIGURE 5 | Density of US-based participants enrolled in each state. There are no states with enrollment between 26 and 100 participants.

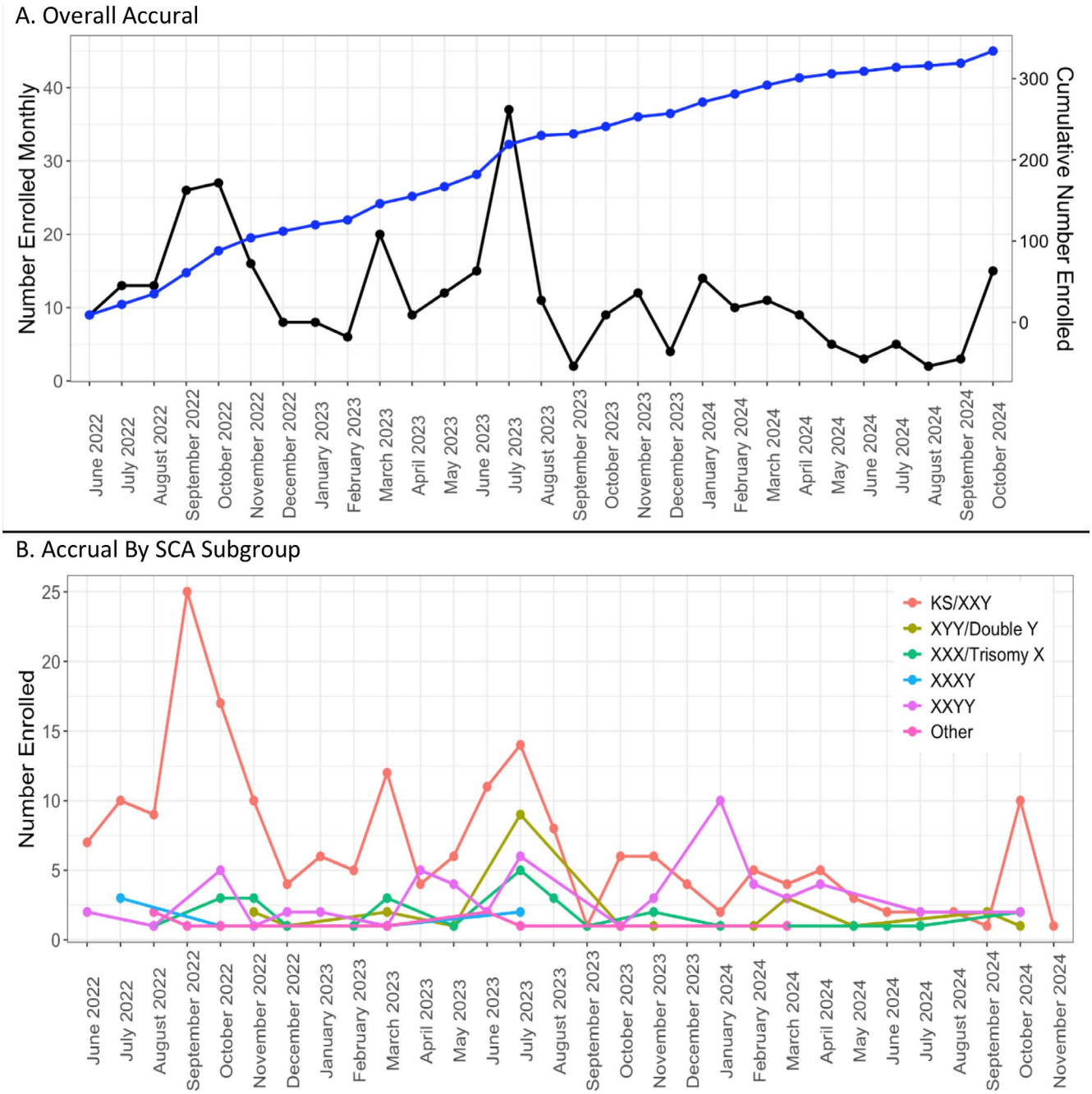

FIGURE 6 | Accrual in the GALAXY registry. (A) Overall accrual. (B) Accrual By SCA subgroup.

for future stratified analyses to determine if these variables affect medical outcomes, development, and quality of life.

The GALAXY Registry is a unique resource for conducting SCA research, offering several strengths that complement historic and current SCA studies. Rare disease registries are recognized as important contributors to evidence-based medicine and are important for describing the natural history and heterogeneity of rare conditions to inform clinical care guidelines (Lee et al. 2011). However, countrywide registries such as those in Denmark (Berglund et al. 2019) and the European Union (Boyd et al. 2011) do not allow for recontact of participants for future studies/surveys, and most individuals in the registries are not prenatally diagnosed; thus, they may be more affected. GALAXY does allow for recontact and has enrolled a sizeable sample of prenatally identified individuals. In contrast, the eXtraordinarY Babies Study (Tartaglia et al. 2020) only recruits individuals who were prenatally identified, providing a unique and important lens for SCA research. However, recruitment is limited in number due to in-depth assessments and the requirement to travel to the study team. GALAXY does not require additional assessments by participants and does not require a prenatal diagnosis, broadening the scope of recruitment. GALAXY and the eXtraordinarY Babies study, as well as many other SCA studies, use common data elements, allowing for integration of data across studies to increase sample size and improve generalizability.

Learning Health Systems, such as PEDSnet (Forrest et al. 2014), also serve an important role in rare disease

TABLE 2 | Diagnostic characteristics, stratified by diagnostic timing.

	Prenatal ($n=139$)	Postnatal ($n=162$)	p
SCA			**<0.001**
47,XXY (KS)[a]	95 (68.3%)	87 (53.7%)	
47,XYY	11 (7.9%)	11 (6.8%)	
47,XXX (Trisomy X)	13 (9.4%)	14 (8.6%)	
47,XXYY	11 (7.9%)	42 (25.9%)	
48,XXXY	2 (1.4%)	5 (3.1%)	
Other[b]	7 (5.0%)	3 (1.9%)	
Genetic test type			
Karyotype	118 (84.9%)	104 (64.2%)	**<0.001**
Microarray	24 (17.3%)	65 (40.1%)	**<0.001**
FISH	32 (23.0%)	7 (4.3%)	**<0.001**
Gene sequencing	1 (0.7%)	9 (5.6%)	**0.023**
Other	4 (2.9%)	2 (1.2%)	0.420
Mosaic	9 (6.5%)	5 (3.1%)	0.181
Reason for genetic testing (dichotomized)			**<0.001**
Incidental[c]	96 (69.1%)	6 (3.7%)	
Secondary to symptoms	43 (30.9%)	156 (96.3%)	

[a]Includes individuals with 47,XX/46,XY.
[b]Includes individuals with mosaic 45X/46XY.
[c]Incidental in the prenatal group was classified as those being tested due to either elective or gender, advanced maternal age, or prior pregnancy or family history of other genetic conditions. Secondary to symptoms was those with abnormal lab or ultrasound results prompting genetic testing.

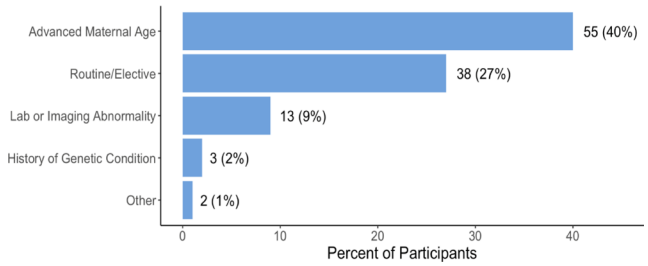

FIGURE 7 | Reasons for prenatal diagnoses. Of the 139 prenatal SCA diagnoses, 55 (40%) were due to advanced maternal age, defined as maternal age at or above 35 at the time of delivery. 38 (27%) were due to routine or elective reasons, including knowing the sex of the fetus; 13 (9%) were due to abnormal lab or ultrasound results; 3 (2%) were due to having a prior pregnancy or family history of genetic conditions; and 2 (1%) were due to other reasons. The remaining 21% had unknown or undocumented indications for prenatal genetic testing.

research by providing clinically verified data; however, it does not include genetic test results as GALAXY does, and the information is limited to what is available within the PEDSnet institution. As GALAXY participants can fill out a Release of Records for multiple institutions, the registry has the potential to acquire all medical records from a participant even across multiple institutions. In contrast, broadly dispersed surveys are common in SCA research, but do not have the ability to confirm SCA diagnosis or other medical history beyond self-report. Data obtained directly from the EHR ensure that the database contains clinically validated data, promoting the reliability of results. GALAXY has the capacity to combine clinically validated information with self-report of SCA confirmed individuals.

While the GALAXY Registry is in its infancy, we have already had a few key observations that have implications for the interpretation of other SCA studies. First, there are marked differences in demographics between the pre- and postnatally diagnosed individuals in this cohort. Most strikingly, 78% of those with a prenatal diagnosis had private insurance, much higher than those with a postnatal diagnosis. Potential reasons for this disparity could include differences in access to (Whitehead et al. 2006), coverage of (Grant et al. 2021; Vahanian et al. 2014), and acceptance of cfDNA screening (Armstrong et al. 2012) and subsequent pregnancy decisions following a positive result. In addition, prior to cfDNA screening, prenatal diagnosis of SCAs was largely limited to women of advanced maternal age who qualified for invasive diagnostic testing; mothers of older age are more likely to be of higher socioeconomic status with private insurance (Nicholls-Dempsey et al. 2023). While reasons for insurance disparities require further investigation, this finding is important to consider when conducting any analyses stratified by the diagnostic timing of SCA, as it is unlikely this is unique to our study. There are also emerging differences in race/ethnicity, method of recruitment, and location of enrollment that should be considered depending on the outcome being studied. There is well-documented evidence of racial, ethnic, and socio-economic disparities in healthcare, including genetic testing and diagnosis (Fraiman and Wojcik 2021), utilization of healthcare services (Chen et al. 2016; National Academies of Sciences 2018), experience with said healthcare services (Hamed et al. 2022), survival rates (GBD US Health Disparities Collaborators 2023) and quality of life/self-reported health (Hayes-Larson et al. 2021). We expect that these factors would similarly impact how and when diagnoses of both the SCA and other comorbid conditions are made, as well as the patient's experience with those diagnoses.

In addition to demographic differences between pre- and postnatal diagnosis, there were also differences in SCA subtype and genetic test obtained. While the trisomy conditions were equally split between pre- and postnatal ascertainment, 80% of participants with tetrasomy conditions were diagnosed postnatally. This is likely a surrogate for phenotypic severity. Previous studies in XYY and Trisomy X syndromes have shown postnatal diagnosis is associated with more neurodevelopmental impairment and medical complexity (Bardsley et al. 2013; Wigby et al. 2016), reflecting the high proportion who do not ever receive a clinical diagnosis. A future direction with the expansion of available clinical data within GALAXY is to quantify the impact of

FIGURE 8 | Reasons for postnatal diagnoses. Of the 162 postnatal SCA diagnoses, the most common indications for genetic testing were developmental delays, learning or behavioral concerns, and dysmorphic features. *excludes autism, ^included any congenital anomaly *other than* genital, heart, and eye.

diagnostic timing on phenotypic severity. Finally, those postnatally diagnosed were more likely to have had a molecular test (e.g., microarray, sequencing) rather than a cytogenetic test (e.g., karyotype, FISH), reflecting current recommendations for genetic testing (Dungan et al. 2023) in the context of multiple congenital anomalies, global developmental delay, autism, epilepsy, and other clinical features. A potential clinical implication of this finding is that individuals who are prenatally identified to have an SCA condition may still warrant a molecular test if they later present with indications for genetic testing, especially if the clinical presentation is not commonly associated with SCA.

GALAXY plans to expand to additional clinical sites; four are currently onboarding, and more additional sites associated with the Association for X and Y Variations (AXYS) Clinic and Research Consortium (ACRC) have expressed interest. Our in-clinic sample is more diverse than participants enrolling remotely, indicating that online recruitment is likely reaching individuals with easier access to and comfort with medical research. Expanding to additional clinical sites will be important for diversifying the participant sample, including more non-XXY individuals and a greater age range and racial composition. In addition, targeting historically underrepresented groups will be undertaken through partnering with community efforts, education about research, and prioritizing transparency. As the GALAXY Registry grows, we anticipate that it will serve as a powerful resource for statistical analyses aimed at clarifying the natural history of SCA and defining predictors of individual-level outcomes.

4.1 | Limitations

The initial limitation is that enrollment is contingent on review of a genetic test result, which excludes individuals who have not received a clinical diagnosis (at least half of all individuals with an SCA) (Bojesen et al. 2003) and also presents a barrier for those with a diagnosis who do not have access to prior results. The current sample is heavily weighted toward younger participants with 47,XXY, which minimizes the utility of the current dataset to answer questions about aging or inform rarer SCA conditions until our cohort expands. Next, the EHR, even when verified by clinicians, may still contain errors of omission or commission that may impact the quality of the data. Similarly,

clinical records are not always available, and not all variables of interest are clinically documented. Finally, the time and financial investment for data abstraction and validation, especially for participants not seen at an affiliated institution, challenges the completeness and sustainability of the registry. Despite these limitations, GALAXY represents an important resource for future SCA research as it is the first and largest clinically validated registry for this population.

5 | Conclusions

The GALAXY Registry—a clinical research registry for individuals with SCAs—was developed with multiple stakeholder perspectives and is diverse and growing. GALAXY provides opportunities for the SCA community to participate in clinical research regardless of their geographic location or ability to access a specialty clinic. GALAXY is building the infrastructure to serve as a resource to connect participants and researchers in the SCA community through mutual partnership and increased opportunities for research. Ultimately, GALAXY aims to provide a conduit for future longitudinal natural history research and clinical trials that will improve care and outcomes for individuals with SCAs.

Author Contributions

Alexandra Carl: conceptualization, data curation, methodology, project administration, and original writing, editing. **Samantha Bothwell:** formal analysis, original writing, editing. **Karli Swenson:** methodology, project administration, editing. **Ryan Bregante:** conceptualization, methodology, funding acquisition, editing. **Lilian Cohen:** conceptualization, methodology, editing. **Virginia Cover:** conceptualization, methodology, funding acquisition, editing. **Anna Dawczyk:** conceptualization, methodology, editing. **Gail Decker:** conceptualization, methodology, funding acquisition, editing. **Stephen B. Gerken:** conceptualization, methodology, editing. **David Hong:** conceptualization, methodology, editing. **Susan Howell:** conceptualization, methodology, editing. **Armin Raznahan:** conceptualization, methodology, editing. **Alan D. Rogol:** conceptualization, methodology, editing. **Nicole Tartaglia:** conceptualization, methodology, editing. **Shanlee Davis:** conceptualization, methodology, funding acquisition, project administration, data curation, editing.

Ethics Statement

This study was approved by the Colorado Multiple Institutional Review Board (20-0482) and all participants gave informed consent prior to any study procedures.

References

Abramsky, L., and J. Chapple. 1997. "47,XXY (Klinefelter Syndrome) and 47,XYY: Estimated Rates of and Indication for Postnatal Diagnosis With Implications for Prenatal Counselling." *Prenatal Diagnosis* 17, no. 4: 363–368. https://doi.org/10.1002/(sici)1097-0223(199704)17:4<363::aid-pd79>3.0.co;2-o.

Armstrong, K., M. Putt, C. H. Halbert, et al. 2012. "The Influence of Health Care Policies and Health Care System Distrust on Willingness to Undergo Genetic Testing." *Medical Care* 50, no. 5: 381–387. https://doi.org/10.1097/MLR.0b013e31824d748b.

Bardsley, M. Z., K. Kowal, C. Levy, et al. 2013. "47,XYY Syndrome: Clinical Phenotype and Timing of Ascertainment." *Journal of Pediatrics* 163, no. 4: 1085–1094. https://doi.org/10.1016/j.jpeds.2013.05.037.

Berglund, A., K. Stochholm, and C. H. Gravholt. 2022. "The Comorbidity Landscape of 47,XXX Syndrome: A Nationwide Epidemiologic Study." *Genetics in Medicine* 24, no. 2: 475–487. https://doi.org/10.1016/j.gim.2021.10.012.

Berglund, A., M. H. Viuff, A. Skakkebæk, S. Chang, K. Stochholm, and C. H. Gravholt. 2019. "Changes in the Cohort Composition of Turner Syndrome and Severe Non-Diagnosis of Klinefelter, 47,XXX and 47,XYY Syndrome: A Nationwide Cohort Study." *Orphanet Journal of Rare Diseases* 14, no. 1: 16. https://doi.org/10.1186/s13023-018-0976-2.

Bojesen, A., S. Juul, and C. H. Gravholt. 2003. "Prenatal and Postnatal Prevalence of Klinefelter Syndrome: A National Registry Study." *Journal of Clinical Endocrinology and Metabolism* 88, no. 2: 622–626. https://doi.org/10.1210/jc.2002-021491.

Boyd, P. A., M. Loane, E. Garne, B. Khoshnood, and H. Dolk. 2011. "Sex Chromosome Trisomies in Europe: Prevalence, Prenatal Detection and Outcome of Pregnancy." *European Journal of Human Genetics* 19, no. 2: 231–234. https://doi.org/10.1038/ejhg.2010.148.

Chen, J., A. Vargas-Bustamante, K. Mortensen, and A. N. Ortega. 2016. "Racial and Ethnic Disparities in Health Care Access and Utilization Under the Affordable Care Act." *Medical Care* 54, no. 2: 140–146. https://doi.org/10.1097/mlr.0000000000000467.

Dungan, J. S., S. Klugman, S. Darilek, et al. 2023. "Noninvasive Prenatal Screening (NIPS) for Fetal Chromosome Abnormalities in a General-Risk Population: An Evidence-Based Clinical Guideline of the American College of Medical Genetics and Genomics (ACMG)." *Genetics in Medicine* 25, no. 2: 100336. https://doi.org/10.1016/j.gim.2022.11.004.

Forrest, C. B., P. A. Margolis, L. C. Bailey, et al. 2014. "PEDSnet: A National Pediatric Learning Health System." *Journal of the American Medical Informatics Association* 21, no. 4: 602–606. https://doi.org/10.1136/amiajnl-2014-002743.

Fraiman, Y. S., and M. H. Wojcik. 2021. "The Influence of Social Determinants of Health on the Genetic Diagnostic Odyssey: Who Remains Undiagnosed, Why, and to What Effect?" *Pediatric Research* 89, no. 2: 295–300. https://doi.org/10.1038/s41390-020-01151-5.

GBD US Health Disparities Collaborators. 2023. "Cause-Specific Mortality by County, Race, and Ethnicity in the USA, 2000-19: A Systematic Analysis of Health Disparities." *Lancet* 402, no. 10407: 1065–1082. https://doi.org/10.1016/S0140-6736(23)01088-7.

Grant, P., S. Langlois, L. D. Lynd, J. C. Austin, and A. M. Elliott. 2021. "Out-Of-Pocket and Private Pay in Clinical Genetic Testing: A Scoping Review." *Clinical Genetics* 100, no. 5: 504–521. https://doi.org/10.1111/cge.14006.

Hamed, S., H. Bradby, B. M. Ahlberg, and S. Thapar-Björkert. 2022. "Racism in Healthcare: A Scoping Review." *BMC Public Health* 22, no. 1: 988. https://doi.org/10.1186/s12889-022-13122-y.

Harris, P. A., R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde. 2009. "Research Electronic Data Capture (REDCap)—A Metadata-Driven Methodology and Workflow Process for Providing Translational Research Informatics Support." *Journal of Biomedical Informatics* 42, no. 2: 377–381. https://doi.org/10.1016/j.jbi.2008.08.010.

Hayes-Larson, E., T. M. Mobley, A. Gilmore-Bykovskyi, et al. 2021. "Racial/Ethnic Differences in Health-Related Quality of Life in Persons With and Without Dementia." *Journal of the American Geriatrics Society* 69, no. 3: 629–636. https://doi.org/10.1111/jgs.16908.

Higurashi, M., K. Iijima, and U. Ikeda. 1979. "Chromosome Survey of Newborn Infants in Tokyo: Follow-Up Study for XXY." In *Sex Chromosome Aneuploidy: Prospective Studies on Children*, vol. XV, 161–174. Alan R Liss Inc.

Jørgensen, I. F., F. Russo, A. B. Jensen, et al. 2019. "Comorbidity Landscape of the Danish Patient Population Affected by Chromosome Abnormalities." *Genetics in Medicine* 21, no. 11: 2485–2495. https://doi.org/10.1038/s41436-019-0519-9.

Lee, N. R., K. C. Lopez, E. I. Adeyemi, and J. N. Giedd. 2011. "Chapter Six—Sex Chromosome Aneuploidies: A Window for Examining the Effects of the X and Y Chromosomes on Speech, Language, and Social Development." In *International Review of Research in Developmental Disabilities*, edited by D. J. Fidler, vol. 40, 139–180. Academic Press. https://doi.org/10.1016/B978-0-12-374478-4.00006-X.

Leonard, M. F., J. E. Schowalter, G. Landy, F. H. Ruddle, and H. A. Lubs. 1979. "Chromosomal Abnormalities in the New Haven Newborn Study: A Prospective Study of Development of Children With Sex Chromosome Anomalies." In *Sex Chromosome Aneuploidy: Prospective Studies on Children*, edited by A. Robinson, H. A. Lubs, and D. Bergsma, vol. XV, 115–160. Alan R Liss Inc.

Linden, M. G., B. G. Bender, and A. Robinson. 1995. "Sex Chromosome Tetrasomy and Pentasomy." *Pediatrics* 96, no. 4: 672. https://link-gale-com.proxy.hsl.ucdenver.edu/apps/doc/A17519296/ITOF?u=denison&sid=bookmark-ITOF&xid=b2d038db.

National Academies of Sciences, E.,&Medicine. 2018. *Health-Care Utilization as a Proxy in Disability Determination*. National Academies Press. https://doi.org/10.17226/24969.

Neilson, J., I. Sillesen, A. M. Sorensen, and K. Sorensen. 1979. "Follow-Up Until Age 4 to 8 of 25 Unselected Children With Sex Chromosome Abnormalities Compaired With Sibs and Controls." In *Sex Chromosome Aneuploidy: Prospective Studies on Children*, edited by A. Robinson, H. A. Lubs, and D. Bergsma, vol. XV, 15–74. Alan R Liss Inc.

Nicholls-Dempsey, L., A. Badeghiesh, H. Baghlaf, and M. H. Dahan. 2023. "How Does High Socioeconomic Status Affect Maternal and Neonatal Pregnancy Outcomes? A Population-Based Study Among American Women." *European Journal of Obstetrics & Gynecology and Reproductive Biology: X* 20: 100248. https://doi.org/10.1016/j.eurox.2023.100248.

Nielsen, J., and M. Wohlert. 1990. "Sex Chromosome Abnormalities Found Among 34,910 Newborn Children: Results From a 13-Year Incidence Study in Arhus, Denmark." *Birth Defects Original Article Series* 26, no. 4: 209–223.

Ratcliffe, S. G., D. Axworthy, and A. Ginsborg. 1979. "The Edinburgh Study of Growth and Development in Children With Sex Chromosome Abnormalities." In *Sex Chromosome Aneuploidy: Prospective Studies on Children*, edited by A. Robinson, H. A. Lubs, and D. Bergsma, vol. XV, 243–260. Alan R Liss Inc.

Robinson, A., M. Puck, B. Pennington, J. Borelli, and M. Hudson. 1979. "Abnormalities of the Sex Chromosomes: A Psrospective Study on Randomly Identified Newborns." In *Sex Chromosome Aneuploidy: Prospective Studies on Children*, edited by A. Robinson, H. A. Lubs, and D. Bergsma, vol. XV, 203–242. Alan R Liss Inc.

Sánchez, X. C., S. Montalbano, M. Vaez, et al. 2023. "Associations of Psychiatric Disorders With Sex Chromosome Aneuploidies in the Danish iPSYCH2015 Dataset: A Case-Cohort Study." *Lancet Psychiatry* 10, no. 2: 129–138. https://doi.org/10.1016/s2215-0366(23)00004-4.

Shankar, R., A. Carl, J. R. Law, et al. 2024. "Inspiring New Science to Guide Healthcare in Turner Syndrome: Rationale, Design, and Methods for the InsighTS Registry." *American Journal of Medical Genetics, Part A* 194, no. 2: 311–319. https://doi.org/10.1002/ajmg.a.63441.

Stewart, D. A., C. T. Netley, J. D. Bailey, et al. 1979. "Growth and Development of Children With X and Y Chromosome Aneuploidy: A Prospective Study." In *Sex Chromosome Aneuploidy: Prospective Studies on Children*, edited by A. Robinson, H. A. Lubs, and D. Bergsma, vol. XV, 75–114. Alan R Liss Inc.

Tartaglia, N., S. Howell, S. Davis, et al. 2020. "Early Neurodevelopmental and Medical Profile in Children With Sex Chromosome Trisomies: Background for the Prospective eXtraordinarY Babies Study to Identify Early Risk Factors and Targets for Intervention." *American Journal of Medical Genetics* 184, no. 2: 428–443. https://doi.org/10.1002/ajmg.c.31807.

Vahanian, S. A., M. Baraa Allaf, C. Yeh, M. R. Chavez, W. L. Kinzler, and A. M. Vintzileos. 2014. "Patient Acceptance of Non-Invasive Testing for Fetal Aneuploidy via Cell-Free Fetal DNA." *Journal of Maternal-Fetal & Neonatal Medicine* 27, no. 1: 106–109. https://doi.org/10.3109/14767058.2013.806477.

Velvin, G., T. Hartman, and T. Bathen. 2022. "Patient Involvement in Rare Diseases Research: A Scoping Review of the Literature and Mixed Method Evaluation of Norwegian Researchers' Experiences and Perceptions." *Orphanet Journal of Rare Diseases* 17, no. 1: 212. https://doi.org/10.1186/s13023-022-02357-y.

Whitehead, N. S., S. A. Rasmussen, S. Cox, and S. F. Posner. 2006. "Prevalence and Predictors of Receipt of Prenatal Information About Genetic Screening." *Prenatal Diagnosis* 26, no. 10: 944–950. https://doi.org/10.1002/pd.1532.

Wigby, K., C. D'Epagnier, S. Howell, et al. 2016. "Expanding the Phenotype of Triple X Syndrome: A Comparison of Prenatal Versus Postnatal Diagnosis." *American Journal of Medical Genetics. Part A* 170, no. 11: 2870–2881. https://doi.org/10.1002/ajmg.a.37688.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information.