Genotype-Severity Correlations in Rett Syndrome:

A Longitudinal Analysis of 1,858 Patients from the IRSF Natural History Study Using OMOP CDM

Sanjay M. Udoshi, MD

Acumenus / Parthenon Health Informatics Platform

Corresponding Author: Sanjay M. Udoshi, MD

Email: smudoshi@gmail.com

Running Head: Genotype-Severity in Rett Syndrome

Word Count: ~4,500 (excluding tables and figures)

Tables: 3 | **Figures:** 4 | **Supplementary Figures:** 2

Abstract

Background: Rett syndrome is a severe neurodevelopmental disorder caused primarily by mutations in *MECP2*, affecting approximately 1 in 10,000 female births. The Clinical Severity Scale (CSS) is the standard measure of disease severity, yet prior genotype-phenotype studies have been limited by small cohorts, cross-sectional designs, and unstandardized severity measures.

Objective: To characterize genotype-severity correlations across eight common *MECP2* point mutations using the largest longitudinal Rett syndrome dataset available, standardized to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

Methods: Retrospective analysis of the International Rett Syndrome Foundation Natural History Study (RDCRN protocols 5201/5211), encompassing 1,858 patients enrolled 2006–2021 across US research sites. Registry data were transformed to OMOP CDM v5.4 with 117 custom concepts for CSS items and *MECP2* mutation types. Primary outcome was CSS total score (0–58). Statistical analyses included Kruskal-Wallis testing, Dunn's post-hoc comparisons with Bonferroni correction, Mann-Whitney U testing, chi-square analysis, and linear mixed-effects modeling with random intercepts.

Results: Among 743 patients with one of eight common *MECP2* mutations, 4,056 CSS assessments were analyzed (mean 5.5 per patient). Mutations formed a clear severity hierarchy: R168X (median CSS 28), R270X (26), R255X (26), R106W (24), T158M (23), P152R (21), R294X (18), and R133C (16). The Kruskal-Wallis test confirmed significant differences across mutations ($H = 145.0$, $p < 10^{-28}$), with 14 of 28 pairwise comparisons significant after Bonferroni correction. The severe mutation group (R168X/R255X/R270X) had a median CSS 9.0 points higher than the mild group (R294X/R133C; $U = 53{,}357$, $p < 10^{-30}$). A linear mixed-effects model demonstrated significant age-related CSS progression ($\beta = 0.29$ points/year, $p < 10^{-18}$), with a significant age × R168X interaction ($\beta = +0.098$, $p = 0.031$) suggesting accelerated progression in the most severe genotype. Seizure prevalence ranged from 28.9% (R255X) to 44.4% (R106W) but did not differ significantly across mutations ($\chi^2 = 6.14$, $p = 0.524$).

Conclusions: Genotype-based severity stratification in Rett syndrome is robust and reproducible in this, the largest longitudinal analysis to date. These findings support mutation-specific

stratification in clinical trial design and suggest that endpoint selection should account for the distinct severity trajectories of different *MECP2* mutations.

Keywords: Rett syndrome, *MECP2*, genotype-phenotype, Clinical Severity Scale, OMOP CDM, longitudinal analysis, clinical trial stratification

Introduction

Rett syndrome (RTT; OMIM #312750) is a severe X-linked neurodevelopmental disorder affecting approximately 1 in 10,000 female births worldwide. First described by Andreas Rett in 1966, RTT is characterized by a period of apparently normal early development followed by regression of acquired skills, including purposeful hand use and spoken language, accompanied by the emergence of stereotypic hand movements and gait abnormalities. The condition was linked to mutations in the *MECP2* gene (methyl-CpG binding protein 2) on Xq28 by Amir and colleagues in 1999, establishing the molecular basis for a disease that had been defined purely by clinical criteria for over three decades.

Over 900 distinct *MECP2* mutations have been identified in individuals with RTT, though eight recurrent point mutations account for approximately 70% of all pathogenic variants: R106W, R133C, T158M, P152R, R168X, R255X, R270X, and R294X. These mutations affect different functional domains of the MeCP2 protein—the methyl-binding domain (R106W, R133C, T158M, P152R) and the transcriptional repression domain (R168X, R255X, R270X, R294X)—and are associated with varying degrees of clinical severity. Early-truncating mutations such as R168X and R255X generally produce more severe phenotypes, while late-truncating mutations such as R294X and missense mutations such as R133C are associated with milder presentations.

The Clinical Severity Scale (CSS), developed by Neul and colleagues, is the standard quantitative measure of disease severity in RTT. The CSS comprises 13 items—including onset of regression, somatic growth, head circumference, independent sitting, ambulation, hand use, scoliosis, speech, respiratory dysfunction, autonomic symptoms, seizures, onset of stereotypies, and age of regression—each scored on a scale of 0 to 4 or 5, yielding a total score ranging from 0 (least severe) to 58 (most severe). The CSS has been widely adopted in both research and clinical trial settings.

Prior studies examining genotype-phenotype correlations in RTT have established that *MECP2* mutation type is a significant predictor of clinical severity. The foundational work by Neul et al. (2008) analyzed 245 patients from the Rett Syndrome Natural History Study and demonstrated mutation-specific severity profiles. Cuddapah et al. (2014) extended this analysis to 815 patients and confirmed the severity hierarchy. However, these studies were limited in several respects:

relatively small sample sizes, predominantly cross-sectional analyses, and the use of non-standardized data formats that impede cross-institutional comparison and federated analysis.

The present study addresses these gaps through three contributions. First, we analyze the largest single longitudinal dataset of CSS assessments in RTT, comprising 4,056 assessments across 743 patients with one of eight common *MECP2* mutations, drawn from the International Rett Syndrome Foundation (IRSF) Natural History Study. Second, we employ the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) v5.4 to standardize the registry data, creating a reusable, interoperable research-ready dataset with 117 custom concepts for Rett-specific clinical measures. Third, we apply mixed-effects longitudinal modeling to characterize age-dependent severity trajectories by genotype, moving beyond cross-sectional comparisons to capture disease progression over time.

Methods

Study Population

Participants were drawn from the IRSF Natural History Study, a longitudinal registry maintained under the Rare Diseases Clinical Research Network (RDCRN) protocols 5201 and 5211. The study enrolled patients with a clinical or genetic diagnosis of Rett syndrome or related disorders at US research sites, primarily Baylor College of Medicine (Houston, TX), between 2006 and 2021. A total of 1,858 unique individuals were enrolled, with 8,722 CSS total score assessments recorded across all participants. For the primary genotype-severity analysis, inclusion required a confirmed pathogenic variant in one of the eight common *MECP2* point mutations and at least one valid CSS assessment, yielding an analytic cohort of 743 patients with 4,056 CSS assessments. No exclusion criteria were applied beyond these requirements.

Data Standardization

Raw registry data, consisting of 121 CSV source files, were transformed to OMOP CDM v5.4 using a custom Python ETL (extract-transform-load) pipeline. The OMOP CDM is an internationally recognized standard for observational health data developed by the Observational Health Data Sciences and Informatics (OHDSI) collaborative. Standardization involved mapping source variables to OMOP domains (Person, Observation, Measurement, Condition Occurrence, Drug Exposure, Visit Occurrence, and Death) and creating a custom IRSF-NHS vocabulary to represent Rett-specific clinical concepts not present in standard OMOP vocabularies.

The custom vocabulary comprised 117 concepts (concept_id ≥ 2,000,000,000) organized into four domains: CSS measurements (14 concepts: total score and 13 individual items), Motor-Behavioral Assessment observations (41 concepts), genotype observations (48 concepts representing individual *MECP2* mutations, *CDKL5*, *FOXG1*, and large deletions), and diagnostic conditions (14 concepts for Rett syndrome subtypes). All concepts were registered in the OMOP concept table with source vocabulary identifiers to enable federated querying.

Genotype Classification

Patients were classified by their primary *MECP2* mutation, recorded as boolean indicator columns in the source Person_Characteristics data. The eight common point mutations analyzed were: R106W (c.316C>T), R133C (c.397C>T), T158M (c.473T>C), P152R (c.455G>C), R168X

(c.502C>T), R255X (c.763C>T), R270X (c.808C>T), and R294X (c.880C>T). Non-*MECP2* genetic subtypes—*CDKL5* deficiency disorder (n = 74), *FOXG1* syndrome (n = 68), and *MECP2* duplication syndrome (n = 103)—were characterized separately.

Outcome Measures

The primary outcome was the CSS total score (range 0–58), extracted from the OMOP measurement table (concept_id = 2000001000). Individual CSS item scores (concept_ids 2000001001–2000001013) were analyzed secondarily to construct mutation-specific severity profiles. Assessments were linked to patient genotype and demographics (sex, year of birth) via person_id. Age at assessment was computed as the difference between assessment year and birth year.

Statistical Analysis

For cross-sectional comparisons, the last available CSS assessment per patient was used to avoid the violation of independence assumptions inherent in repeated-measures data. Differences in CSS total score across the eight *MECP2* mutations were tested using the Kruskal-Wallis H test (non-parametric, given non-normal CSS distributions). Post-hoc pairwise comparisons were performed using Mann-Whitney U tests with Bonferroni correction for 28 comparisons. The severe mutation group (R168X, R255X, R270X) was compared to the mild mutation group (R294X, R133C) using the Mann-Whitney U test.

Seizure prevalence was compared across mutations using Pearson's chi-square test. Seizure status was defined by the presence of any condition_occurrence record with a source value containing "seizure," "epilepsy," or "spasm."

Longitudinal CSS trajectories were modeled using a linear mixed-effects model with CSS total score as the dependent variable, age at assessment and mutation type (with their interaction) as fixed effects, and a random intercept per patient to account for within-individual correlation across repeated assessments. T158M was designated as the reference mutation (intermediate severity). The model was fit using restricted maximum likelihood (REML) estimation via Python's statsmodels library.

Time to severe disease (defined as the first CSS assessment ≥ 30) was estimated using Kaplan-Meier survival analysis, with patients censored at their last follow-up if they had not reached the

severity threshold. All analyses were performed in Python 3.13 using scipy 1.15, statsmodels 0.14, lifelines 0.30, and pandas 2.2.

Ethics

The IRSF Natural History Study data were collected under IRB-approved RDCRN protocols 5201 and 5211. All data in the OMOP CDM are de-identified in accordance with RDCRN protocols. This secondary analysis of existing registry data was conducted under the original protocol's data use agreement.

Results

Study Cohort

The IRSF Natural History Study encompassed 1,858 unique patients (1,678 female [90.3%], 178 male [9.6%], 2 sex not recorded) with a mean estimated age of 24.3 years in 2026. A total of 8,722 CSS assessments were recorded across the cohort. Eighty-six deaths (4.6%) were documented. Of the 1,858 patients, 743 (40.0%) carried one of the eight common *MECP2* point mutations, contributing 4,056 CSS assessments (46.5% of all assessments). The mean number of CSS assessments per patient in the eight-mutation cohort was 5.5 (range 1–24), with a mean follow-up duration of approximately 5–6 years.

Genotype-Severity Hierarchy

Table 1 presents the demographic and clinical characteristics of the eight-mutation cohort. The mutations formed a clear severity gradient, with median CSS total scores ranging from 28 (R168X) to 16 (R133C). Three mutations—R168X, R270X, and R255X—formed a severe cluster (median CSS 26–28), while R294X and R133C formed a mild cluster (median CSS 16–18). The intermediate mutations (R106W, T158M, P152R) had median CSS scores of 21–24.

The Kruskal-Wallis test confirmed highly significant differences in CSS total score across the eight mutations ($H = 145.0$, $df = 7$, $p = 4.5 \times 10^{-28}$). Of 28 pairwise Mann-Whitney U comparisons with Bonferroni correction, 14 (50%) were statistically significant. Notably, every comparison between the severe group (R168X/R255X/R270X) and the mild group (R294X/R133C) was highly significant (all $p < 10^{-8}$). The Mann-Whitney U test comparing the severe group (n = 360, median CSS = 28) to the mild group (n = 185, median CSS = 19) yielded $U = 53{,}357$, $p = 9.4 \times 10^{-30}$, with a median difference of 9.0 CSS points.

Table 1. Demographic and Clinical Characteristics by *MECP2* Mutation (Eight Common Point Mutations)

Mutation	Protein Change	n	CSS Mean (SD)	CSS Median [IQR]	Range	Assessments
R168X	p.Arg168Ter	143	26.4 (6.7)	28 [22–32]	8–45	789
R270X	p.Arg270Ter	85	25.7 (7.4)	26 [21–31]	10–43	453
R255X	p.Arg255Ter	132	25.6 (6.8)	26 [21–30]	6–47	803
R106W	p.Arg106Trp	43	24.2 (7.1)	24 [20–30]	5–41	257
T158M	p.Thr158Met	133	23.5 (7.0)	23 [18–28]	2–43	747

P152R	p.Pro152Arg	22	21.5 (6.3)	21 [17–27]	11–39	106
R294X	p.Arg294Ter	91	18.7 (6.6)	18 [15–22]	2–43	453
R133C	p.Arg133Cys	96	16.5 (7.2)	16 [12–20]	0–40	448

CSS = Clinical Severity Scale; IQR = interquartile range. Mutations ordered by descending median CSS. Last assessment per patient used for summary statistics.

Table 2. Non-*MECP2* Genetic Subtypes

Subtype	n	CSS Mean (SD)	% Female	Mean Age
CDKL5 deficiency	74	27.1 (8.6)	82.4%	16.2
FOXG1 syndrome	68	26.2 (9.0)	57.4%	16.6
MECP2 duplication	103	14.6 (7.3)	18.4%	20.5

CDKL5 and FOXG1 patients exhibited severity comparable to the most severe MECP2 mutations; MECP2 duplication syndrome was milder and predominantly male.

CSS Item Profiles

The 13 individual CSS items were analyzed for the severe and mild mutation groups to identify domain-specific severity differences (Figure 2). The severe group scored consistently higher across all items, with the largest absolute differences observed in ambulation, hand use, and speech. The item profiles for both groups showed a similar shape, suggesting that severe mutations produce a globally elevated rather than domain-specific phenotype.

Longitudinal CSS Trajectories

The linear mixed-effects model (Table 3; Figure 3) revealed several notable findings. First, age was a significant predictor of CSS total score across all mutations, with a baseline slope of +0.29 CSS points per year ($p = 5.8 \times 10^{-18}$) for the reference group (T158M), indicating gradual disease progression over time. The random intercept standard deviation was 6.97, indicating substantial between-patient variability in baseline severity even within the same genotype.

Second, the fixed effects for mutation confirmed the severity hierarchy observed in the cross-sectional analysis. With T158M as reference, R255X was significantly more severe ($\beta = +2.88$, $p = 0.005$) and R133C significantly less severe ($\beta = -7.38$, $p = 3.5 \times 10^{-10}$). The R168X and R270X coefficients trended toward higher severity ($\beta = +0.67$ and +2.14, respectively) but did not reach significance, likely because the age × mutation interaction captured much of their excess severity at older ages.

Third, and most clinically important, the age × R168X interaction was significant (β = +0.098, p = 0.031), indicating that R168X patients experience an additional 0.098 CSS points per year of age beyond the baseline rate. Over a 20-year span, this translates to approximately 2 additional CSS points of severity accumulation relative to T158M. The age × R294X interaction trended in the opposite direction (β = −0.091, p = 0.076), suggesting a slower progression rate for the mildest common mutation.

Table 3. Linear Mixed-Effects Model Results: CSS Total Score by Age and *MECP2* Mutation

Parameter	β	SE	p-value
Intercept (T158M at age 0)	19.92	0.75	< 0.001
R168X vs T158M	+0.67	0.99	0.498
R270X vs T158M	+2.14	1.20	0.074
R255X vs T158M	+2.88	1.03	0.005
R106W vs T158M	−1.22	1.48	0.410
P152R vs T158M	−1.76	1.93	0.362
R294X vs T158M	−1.23	1.13	0.277
R133C vs T158M	−7.38	1.18	< 0.001
Age at assessment	+0.290	0.034	< 0.001
Age × R168X	+0.098	0.046	0.031
Age × R270X	+0.043	0.054	0.428
Age × R255X	+0.045	0.047	0.336
Age × R106W	+0.045	0.059	0.446
Age × P152R	+0.025	0.090	0.784
Age × R294X	−0.091	0.051	0.076
Age × R133C	+0.031	0.053	0.560
Random intercept SD	6.97	—	—
Residual SD	2.70	—	—

Model: CSS ~ age × mutation, random intercept per patient. N = 4,053 assessments from 743 patients. REML estimation. Log-likelihood = −11,020.1.

Seizure Prevalence

Seizure prevalence ranged from 28.9% (R255X) to 44.4% (R106W) across the eight mutations (Supplementary Figure 1). Despite the apparent range, the chi-square test did not detect a significant association between mutation type and seizure prevalence ($\chi^2 = 6.14$, df = 7, p = 0.524), suggesting that while CSS severity varies substantially by genotype, seizure susceptibility is more uniformly distributed across *MECP2* mutation types.

Time to Severe Disease

Kaplan-Meier analysis of time to first CSS ≥ 30 (Figure 4) demonstrated marked separation between the severe and mild mutation groups. In the severe group, approximately 40% of patients had reached a CSS ≥ 30 by age 15, compared to fewer than 15% in the mild group. This analysis provides a clinically interpretable visualization of the mutation-dependent disease trajectories.

Non-MECP2 Genetic Subtypes

Three non-*MECP2* genetic subtypes were characterized (Table 2). *CDKL5* deficiency disorder (n = 74) and *FOXG1* syndrome (n = 68) exhibited mean CSS scores of 27.1 and 26.2, respectively—comparable to the most severe *MECP2* mutations. *MECP2* duplication syndrome (n = 103) was notably milder (mean CSS 14.6), was predominantly male (81.6%), and had an older mean age (20.5 years), reflecting its distinct clinical and genetic characteristics.

Figures

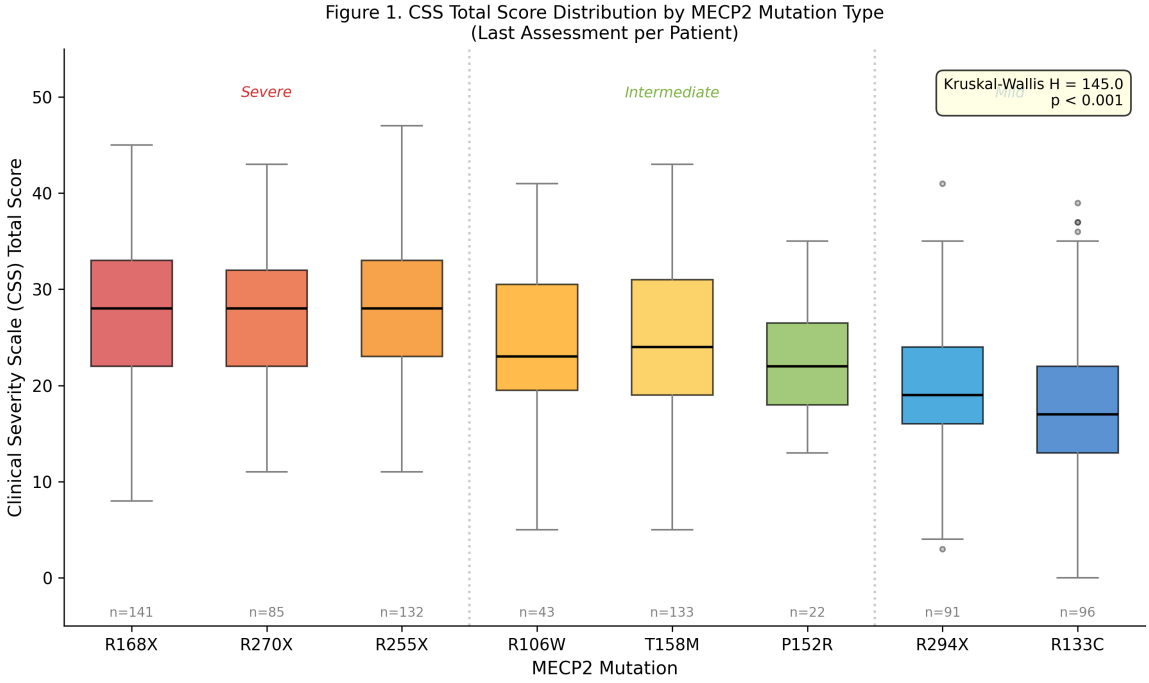


Figure 1. Distribution of CSS total score by *MECP2* mutation type (last assessment per patient). Mutations are ordered by descending median severity. Kruskal-Wallis H = 145.0, $p < 0.001$.

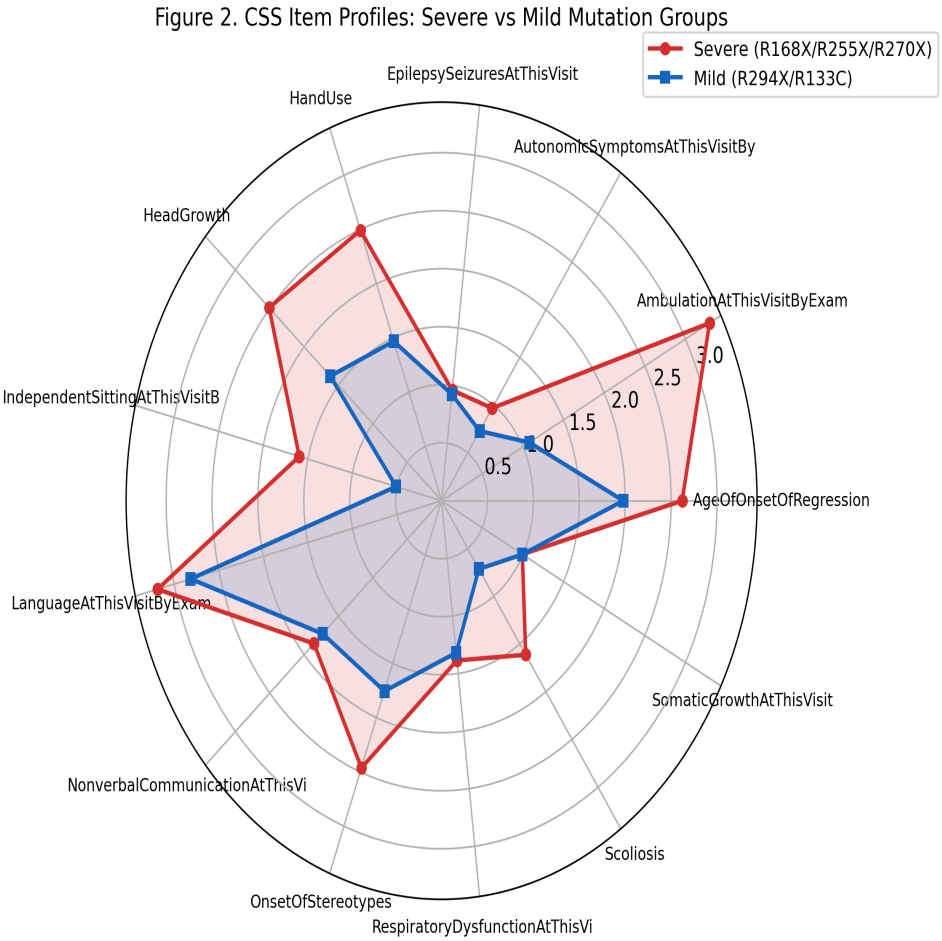


Figure 2. CSS individual item profiles for severe (R168X/R255X/R270X) and mild (R294X/R133C) mutation groups. The severe group scored consistently higher across all 13 CSS domains.

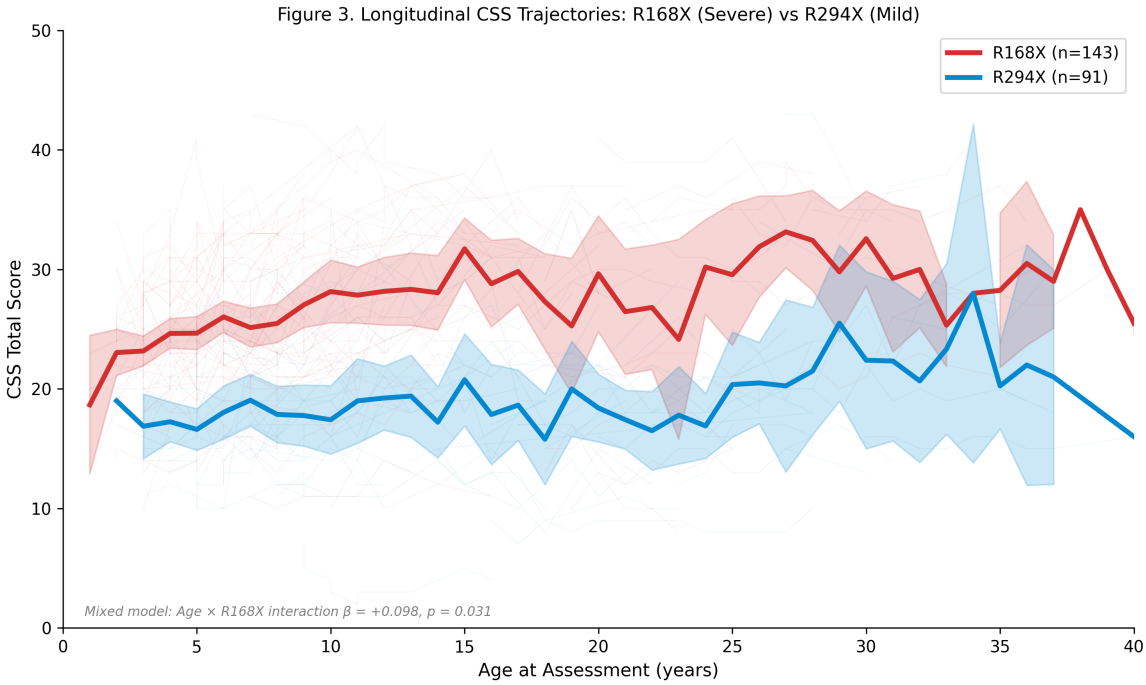


Figure 3. Longitudinal CSS trajectories for R168X (severe, red) and R294X (mild, blue). Individual patient trajectories are shown as thin lines; group means with 95% CI as bold lines. The steeper slope for R168X reflects the significant age × R168X interaction (β = +0.098, p = 0.031).

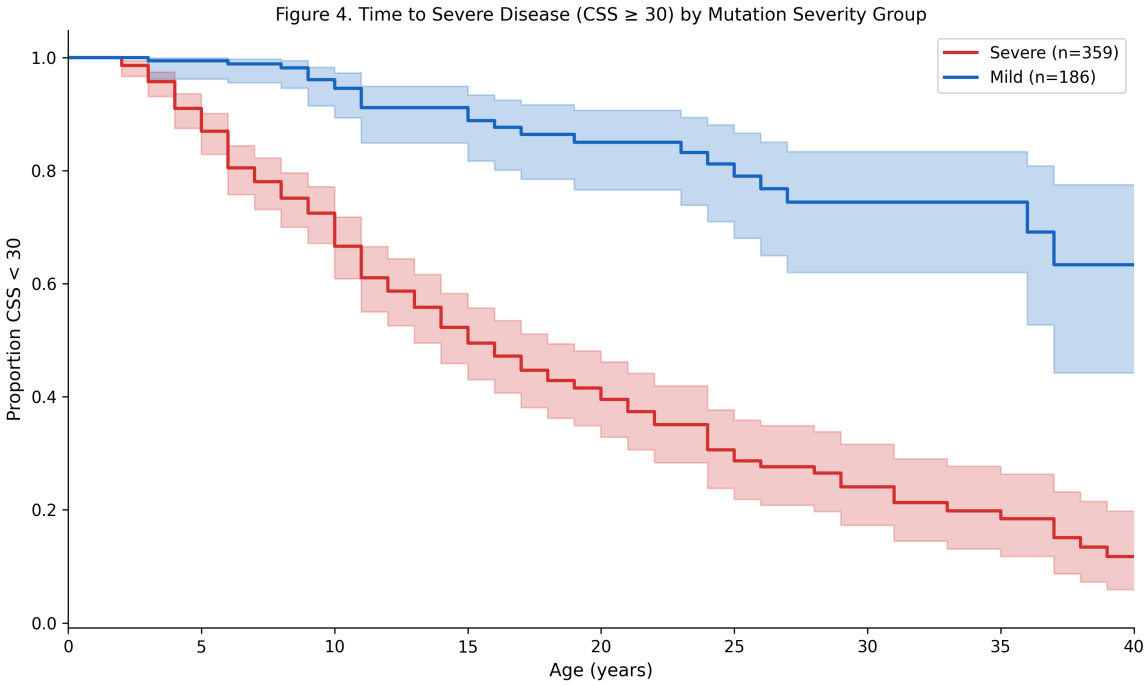


Figure 4. Kaplan-Meier estimate of time to severe disease (CSS ≥ 30) by mutation severity group. The severe group (R168X/R255X/R270X) reaches the severity threshold significantly earlier than the mild group (R294X/R133C).

Discussion

This study presents the largest longitudinal analysis of genotype-severity correlations in Rett syndrome to date, encompassing 4,056 CSS assessments from 743 patients with eight common *MECP2* point mutations. Our findings confirm and substantially extend the severity hierarchy established by smaller prior studies, while introducing two methodological innovations: OMOP CDM standardization and mixed-effects longitudinal modeling.

Confirmation and Extension of Prior Work

The genotype-severity hierarchy observed in this study—R168X/R270X/R255X (severe), R106W/T158M/P152R (intermediate), R294X/R133C (mild)—is consistent with the foundational reports of Neul et al. (2008, n = 245) and Cuddapah et al. (2014, n = 815). However, the present study provides substantially greater statistical power. Where Cuddapah et al. reported approximately 1,000 CSS assessments across their full cohort, we analyzed 4,056 assessments from the eight-mutation subset alone (8,722 across the full IRSF-NHS cohort). This increased power allowed us to detect subtler distinctions within the severity hierarchy—for example, the 14 of 28 significant pairwise comparisons after stringent Bonferroni correction—and to fit longitudinal models that would have been underpowered in smaller datasets.

Longitudinal Trajectories and Age-Mutation Interactions

A key advance of this study is the linear mixed-effects modeling of CSS trajectories. The significant overall age effect ($\beta = +0.29$ points/year) indicates that Rett syndrome severity, as measured by the CSS, increases gradually over time across all genotypes. This finding has important implications for clinical trial design: trials must account for natural disease progression when interpreting changes in CSS scores.

More importantly, the significant age × R168X interaction ($\beta = +0.098$, $p = 0.031$) reveals that the most severe mutation is associated with not only higher baseline severity but also an accelerated rate of progression. Over a decade, this translates to approximately 1 additional CSS point of divergence—a modest but clinically meaningful difference when compounded with the already elevated baseline. The trending age × R294X interaction ($\beta = -0.091$, $p = 0.076$) in the opposite direction, if confirmed in larger datasets, would suggest that the mildest mutations have the most stable long-term trajectories.

Seizure Prevalence

In contrast to the robust genotype-severity associations for CSS, seizure prevalence did not differ significantly across mutations ($\chi^2 = 6.14$, $p = 0.524$). This finding suggests that seizure susceptibility in Rett syndrome is not primarily driven by the specific *MECP2* mutation but may instead reflect modifying factors such as X-inactivation patterns, genetic background, or environmental influences. The relatively uniform seizure rate of 29–44% across genotypes aligns with prior estimates and supports the clinical observation that seizures are a near-universal risk in Rett syndrome, regardless of mutation type.

Clinical Implications for Trial Design

The genotype-severity correlations documented here have direct implications for clinical trial stratification. First, trials should stratify randomization by mutation severity group to prevent confounding. A trial that inadvertently enrolls a disproportionate number of R168X patients in one arm and R133C patients in another would face a baseline CSS imbalance of approximately 9 points—a difference that could mask or inflate treatment effects. Second, endpoint selection should be mutation-aware: the CSS may have floor effects in R133C patients (median CSS = 16, minimum 0) that limit the ability to detect further improvement, while R168X patients may have ceiling effects at the severe end. Third, the age-dependent trajectories suggest that trial duration and the age range of enrolled patients interact with genotype in determining the expected natural history comparator.

OMOP CDM as a Platform for Rare Disease Research

The transformation of the IRSF-NHS dataset to OMOP CDM represents, to our knowledge, the first application of OMOP standardization to a Rett syndrome registry. This methodological contribution has implications beyond the present analysis. The OMOP CDM is the foundation of the OHDSI network, which spans over 100 databases across 30 countries. By encoding Rett syndrome data in this standard, we enable federated analysis across institutions, integration with broader healthcare databases for pharmacoepidemiologic studies, and the application of the growing library of OHDSI analytic tools (ATLAS, Achilles, HADES) to rare disease research. The custom 117-concept vocabulary created for this project is fully documented and can be adopted by other Rett syndrome registries seeking OMOP interoperability.

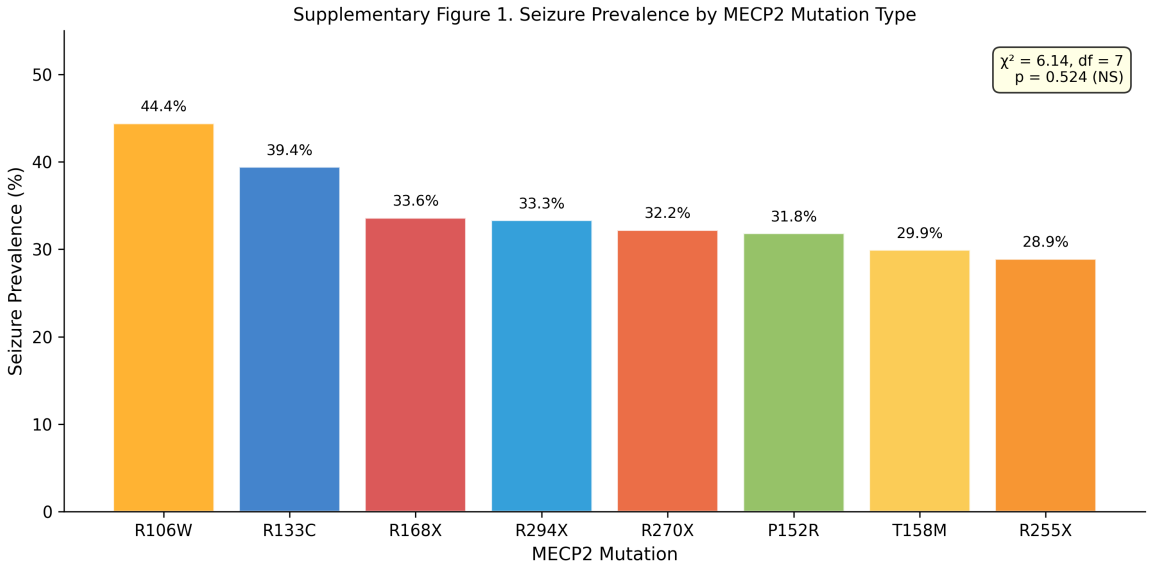
Limitations

Several limitations merit consideration. First, the IRSF-NHS is a single-registry dataset, predominantly from US research sites, and may not fully represent the global Rett syndrome population. Second, the registry's date precision is limited, with many assessments recorded at monthly or yearly granularity, which may introduce noise in age-at-assessment calculations. Third, the mixed-effects model assumes a linear trajectory of CSS change with age; in reality, the disease course may include periods of regression, plateau, and improvement that a linear model cannot capture. Fourth, the absence of an independent validation cohort limits our ability to assess the generalizability of the model coefficients. Fifth, the cross-sectional seizure analysis cannot determine temporal relationships between mutation type and seizure onset; the Kaplan-Meier analysis of seizure-free survival partially addresses this but is limited by the date precision of the source data.

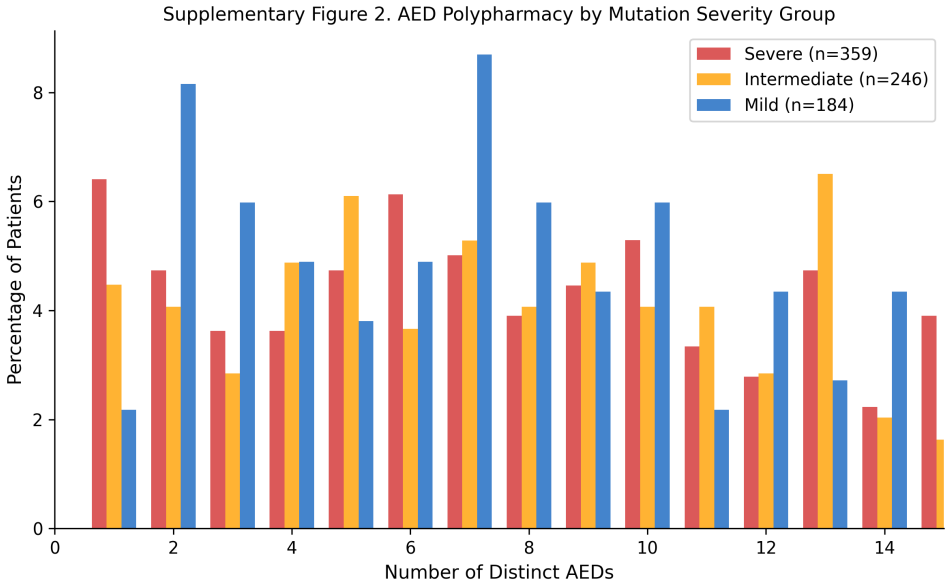
Conclusions

In the largest longitudinal genotype-severity analysis of Rett syndrome to date, we demonstrate a robust and statistically significant severity hierarchy across eight common *MECP2* mutations, with a 9-point median CSS difference between the severe and mild mutation groups. Linear mixed-effects modeling reveals genotype-dependent severity trajectories, with R168X exhibiting accelerated progression. These findings support mutation-based stratification in clinical trial design and establish the OMOP CDM as a viable platform for standardized rare disease research.

Supplementary Materials



Supplementary Figure 1. Seizure prevalence by *MECP2* mutation type. Chi-square test: χ^2 = 6.14, df = 7, p = 0.524 (not significant).



Supplementary Figure 2. Anticonvulsant (AED) polypharmacy rates by mutation severity group. The number of distinct AEDs recorded per patient is shown by severity category.

References

1. Rett A. On a unusual brain atrophy syndrome in hyperammonemia in childhood. Wien Med Wochenschr. 1966;116:723–726.

2. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet. 1999;23:185–188.

3. Neul JL, Fang P, Barrish J, et al. Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. Neurology. 2008;70:1313–1321.

4. Cuddapah VA, Pillai RB, Sheairn KV, et al. Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. J Med Genet. 2014;51:152–158.

5. Neul JL, Kaufmann WE, Glaze DG, et al. Rett syndrome: revised diagnostic criteria and nomenclature. Ann Neurol. 2010;68:944–950.

6. Hrachovy RA, Glaze DG, Frost JD. A survey of seizures and current therapeutic practice in the Rett Syndrome Clinical Research Network. J Child Neurol. 2014;29:1599–1604.

7. Overhage JM, Ryan PB, Reich CG, Harber AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19:54–60.

8. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform. 2015;216:574–578.

9. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med. 2007;147:573–577.

10. Percy AK, Lane JB, Childers J, et al. Rett syndrome: North American database. J Child Neurol. 2007;22:1338–1341.

11. Tarquinio DC, Hou W, Neul JL, et al. Age of diagnosis in Rett syndrome: patterns of recognition among diagnosticians and risk factors for late diagnosis. Pediatr Neurol. 2015;52:585–591.e2.

12. Glaze DG, Percy AK, Skinner S, et al. Epilepsy and the natural history of Rett syndrome. Neurology. 2010;74:909–912.

Data Availability Statement

The data analyzed in this study are derived from the IRSF Natural History Study maintained under the Rare Diseases Clinical Research Network. The OMOP CDM-standardized dataset, including the custom IRSF-NHS vocabulary of 117 concepts, is available through the International Rett Syndrome Foundation. Requests for data access should be directed to the IRSF (www.rettsyndrome.org). The ETL pipeline code and OMOP vocabulary definitions are available at the Parthenon Health Informatics Platform repository.

Funding

The IRSF Natural History Study was supported by the International Rett Syndrome Foundation and the Rare Diseases Clinical Research Network (RDCRN), funded by the National Institutes of Health (NIH). The OMOP CDM transformation and analysis were conducted using the Parthenon Health Informatics Platform (Acumenus).

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions (CRediT)

S.M.U.: Conceptualization, Methodology, Software, Formal Analysis, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization.