Establishing the Rett Syndrome Developmental Milestone Assessment (RS-DMA) as a Primary Endpoint for Interventional Studies

Poster 7

Megan Quarmley,[1] Minna Montgomery,[2] Christopher Jones,[1] Robin C. Ryther,[3] David Lieberman,[4] Elizabeth Berry-Kravis,[5] Kristin Phillips,[2] Emily McGinnis,[2] Rumana Haque-Ahmed,[2] Suku Nagendran,[2] Tessa Clarkson[1]

[1]Psychlomere, Littleton, CO, USA; [2]Taysha Gene Therapies, Dallas, TX, USA; [3]Washington University School of Medicine in St. Louis, St. Louis, MO, USA; [4]Boston Children's Hospital, Boston, MA, USA; [5]Rush University Medical Center, Chicago, IL, USA

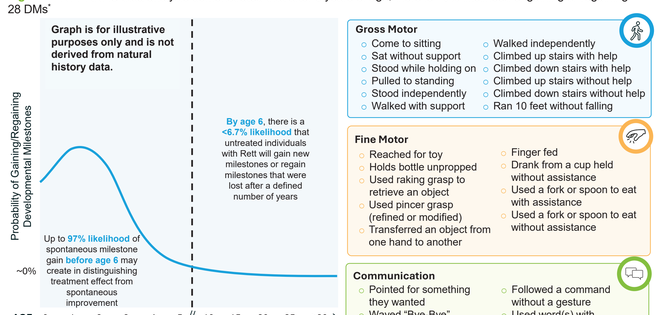
Aims

To psychometrically validate the RS-DMA — a standardized video-based, individualized assessment evaluating the gain or regain of 28 developmental milestones across communication, fine motor, and gross motor domains — as a fit-for-purpose primary endpoint for use in RTT interventional trials, including in a single-arm pivotal study design using a plateau-anchored, baseline-adjusted (PABA) analysis framework.

Background

- FDA's 2025 draft guidance supports evaluating clinical benefit in single-arm designs using alternative controls such as our novel Plateau-Anchored Baseline-Adjusted (PABA) analysis framework for gene therapies in heterogenous rare diseases[1–3]
- Endpoints for such trial designs must be fit-for-purpose, psychometrically strong, control for type 1 errors, be clinically meaningful, incorporate patient and caregiver priorities, and measure individualized improvements for rigorous, blinded evaluation of investigational therapies
- Rett syndrome (RTT) is a rare neurodevelopmental disorder characterized by lost or missed developmental milestones (DMs) across three functional domains (communication, fine motor, gross motor function) impacting activities of daily living,[4] with substantial clinical heterogeneity and is the focus of multiple ongoing gene therapy clinical trials leveraging single-arm designs
- Existing Clinical Outcome Assessments (COAs) in RTT are not well-suited as primary endpoints in single-arm designs using a PABA analysis framework[5-8] as they cannot be bias-mitigated through blinded video review, lack a natural history anchored plateau, or rely on summary scores that lack sensitivity to individualized change
- Recent analyses of the U.S.-based RTT and RTT-related Disorders Natural History Study (RNHS) identified 28 developmental milestones across communication, fine motor function, and gross motor domains that establish a developmental plateau that could be used to design a novel individualized endpoint. These milestones are statistically unlikely to be gained or regained (after a prespecified time since loss) in the plateau population after a prespecified time since loss (Figure 1)[9] (see poster #39)
- Building on this work, we developed the Rett Syndrome Developmental Milestones Assessment (RS-DMA) as a regulatory-aligned standardized primary endpoint suitable for use in a Phase 3 interventional trial that measures individualized change in milestone attainment and identifies treatment responders in an PABA aggregate responder analysis compared to RNHS natural history data
- Here we psychometrically validated the RS-DMA for use in classic RTT in a non-interventional study, RESUME.

Figure 1: RTT RNHS data analysis demonstrated that after 6 years of age, there is ~0% likelihood of gaining or regaining 28 DMs*



Gross Motor
- Come to sitting
- Sat without support
- Stood while holding on
- Pulled to standing
- Stood independently
- Walked with support
- Walked independently
- Climbed up stairs with help
- Climbed down stairs with help
- Climbed up stairs without help
- Climbed down stairs without help
- Ran 10 feet without falling

Fine Motor
- Reached for toy
- Holds bottle unpropped
- Used raking grasp to retrieve an object
- Used pincer grasp (refined or modified)
- Transferred an object from one hand to another
- Finger fed
- Drank from a cup held without assistance
- Used a fork or spoon to eat with assistance
- Used a fork or spoon to eat without assistance

Communication
- Pointed for something they wanted
- Waved "Bye-Bye"
- Followed a command with a gesture
- Identified body parts (pointed with eyes or fingers)
- Followed a command without a gesture
- Used word(s) with meaning
- Spoke in phrases (2 words or more) with meaning

Methods

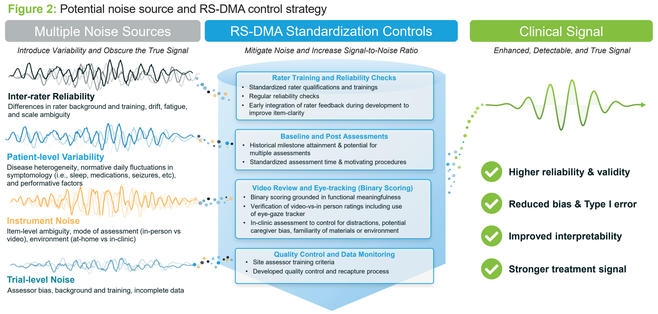
Study Design

- The RESUME study is a U.S. based, non-interventional, and multisite longitudinal study designed to establish trial-ready video-capture and assessment methodologies in RTT
- Thirteen participants (N=13; Age: M=11.3, SD=4.7; CGI-S: M=4.9, SD=0.9) completed the RS-DMA, alongside other assessments, to evaluate the psychometrics and trial-readiness

RS-DMA Construct Design

- In small-sample studies, strong psychometrics and completeness of data are critical for detecting treatment responses
- Designing a tool that achieves strong psychometrics (high signal to noise ratio), requires tight control over potential sources of noise (Figure 2)

Figure 2: Potential noise source and RS-DMA control strategy



Operationalizing Definitions:

- Matched RSNHS milestones with well-established definitions from standardized assessments
- Refined definitions to ensure comparability with the RSNHS interviews
- Dichotomized scoring ("demonstrated" or "not demonstrated") to allow for a responder analysis and improve reliability
- Refined definitions from central rater feedback to ensure strong feasibility and reliability of video-review

Historical Caregiver Interview: *used to determine milestone eligibility prior to treatment based on age and time since loss for each milestone*

- Replicated method used in the RNHS to gather retrospective data on milestone gain, loss, or regain using our RS-DMA construct definitions, including memory-aided cues and medical record checks to mitigate recall bias

Performative RS-DMA In-clinic Assessment: *used to evaluate current functional abilities and potential treatment response via blinded video-based central rating*

- Standardized materials and administration procedures: fixed order, prompts, levels of support and use of assistive devices and modifications (Figure 3A)
- Established video capture procedures with 2 stationary cameras and 1 Pupil Labs eye-tracking glasses with recorded gaze-overlay in real-time (Figure 3B)
- Developed quality control and training procedures to ensure trial-readiness

Figure 3: (A) Kit for RS-DMA enabling standardization of materials and administration procedures; (B) Eye-tracking overlay using Pupil Labs glasses on the Identified body parts with a participant with RTT

In (B), the red circle represents real-time gaze overlay during the identifies body parts item asking to find the dog's mouth, nose, eyes, and hands respectively to ensure video review replicated in-person assessment perspective of eye-gaze responses.

Psychometric Analyses

- We examined the feasibility, fidelity, reliability, convergent validity and clinical meaningfulness of the RS-DMA using R "irr" for kappa analyses and "boot" for bootstrapping confidence intervals and standard deviations
- Clinical meaningfulness of milestones was evaluated using a questionnaire asking caregivers for each milestone endorsed as "not currently able" to consider if gaining or regaining a milestone would be clinically meaningful in response to a hypothetical treatment

Results

Feasibility:

Is it possible to rate milestones using video capture methods?

- The RS-DMA met feasibility criteria, with a mean feasibility score of 8.27 (SD=1.17), indicating 83% of RS-DMA video recordings were deemed feasible for central rating
- Though high, to avoid any missed measurements, quality control and recapture processes can be implemented in interventional settings to avoid any data loss.

Is video rating as accurate as in-person assessment?

- The reliability between Central Raters and Site Assessors Raters were substantial to excellent agreement on 27 of the 28 milestone items (Table 1), and moderate agreement on 1 milestone, "Identifies body parts," with some raters reporting qualitatively that the eye tracking device assisted in their ability to rate this milestone (Figure 3B). Video-based rating facilitated more complete and accurate scoring by enabling raters to pause and review recordings, whereas site assessors scored in real time while also focusing on test administration.

Table 1: Summary of the psychometrics of RS-DMA across the 28 milestones

	Developmental Milestone	Feasibility of Video-Based Rating				Inter Rater Reliability			
		N	% Agree	PABAK	Interpretation	N	% Agree	PABAK	Interpretation
FM	M10: Finger fed	11	100	1	Excellent	8	100	1	Excellent
FM	M11: Pincer grasp	11	81.8	0.636	Substantial	8	89.8	0.797	Substantial
FM	M12: Taken a drink from a cup held without assistance	11	100	1	Excellent	8	96.2	0.923	Excellent
FM	M13: Hold bottle	11	100	1	Excellent	8	91.5	0.83	Excellent
FM	M14: Used a spoon (or fork) to eat without assistance	11	100	1	Excellent	8	94.8	0.896	Excellent
FM	M15: Used a spoon (or fork) to eat with assistance	11	100	1	Excellent	8	96.2	0.923	Excellent
FM	M16: Transferred an object from one hand to the other	11	100	1	Excellent	8	100	1	Excellent
FM	M8: Reach for toy	11	81.8	0.636	Substantial	8	98.1	0.962	Excellent
FM	M9: Used raking grasp to retrieve an object	11	90.9	0.818	Excellent	8	92.9	0.857	Excellent
GM	M17: Come to sitting	11	100	1	Excellent	8	89.8	0.797	Substantial
GM	M18: Sat without Support	11	100	1	Excellent	8	88.7	0.775	Substantial
GM	M19: Pulled to standing	11	81.8	0.636	Substantial	8	92	0.841	Excellent
GM	M20: Stood independently	11	100	1	Excellent	8	96.2	0.923	Excellent
GM	M21: Stood while holding on	11	100	1	Excellent	8	84.3	0.687	Substantial
GM	M22: Walked independently	11	100	1	Excellent	8	95.9	0.918	Excellent
GM	M23: Walked with support	11	100	1	Excellent	8	89.8	0.797	Substantial
GM	M24: Ran 10 ft without falling	11	100	1	Excellent	8	100	1	Excellent
GM	M25: Climbed up stairs without help	11	90.9	0.818	Excellent	8	95.9	0.918	Excellent
GM	M26: Climbed up stairs with help	11	90.9	0.818	Excellent	8	96.7	0.934	Excellent
GM	M27: Climbed downstairs without help	11	90.9	0.818	Excellent	8	94	0.879	Excellent
GM	M28: Climbed downstairs with help	11	90.9	0.818	Excellent	8	96.7	0.934	Excellent
SC	M1: Waves bye-bye	11	100	1	Excellent	8	94.2	0.885	Excellent
SC	M2: Followed a command without a gesture	11	100	1	Excellent	8	92	0.841	Excellent
SC	M3: Followed a command with a gesture	11	81.8	0.636	Substantial	8	83.2	0.665	Substantial
SC	M4: Identified body parts (pointed with eyes or fingers)	11	72.7	0.455	Moderate	8	65.1	0.302	Fair
SC	M5: Used words with meaning	11	100	1	Excellent	8	100	1	Excellent
SC	M6: Spoken in phrases (2 words or more) with meaning	11	100	1	Excellent	8	96.7	0.934	Excellent
SC	M7: Pointed for something they wanted	11	90.9	0.818	Excellent	8	90.1	0.802	Excellent

Fidelity: How consistent are the administrations across sites with the RS-DMA standardization?

- The RS-DMA met fidelity criteria, with a mean score of 85.8% (SD=3.30%). The RS-DMA domains of communication, fine motor, and gross motor domains achieved mean fidelity scores of 86.5% (SD=6.58%), 86.8% (SD=5.37%) and 88.4% (SD=5.91%), respectively. There were no significant differences across sites.
- These findings indicate that administration fidelity was uniformly high across three decentralized study sites following protocol-specified assessor training, standardization, and quality-control procedures, underscoring confidence in the RS-DMA's consistent delivery and trial-readiness

Reliability:

How consistent are central raters in their ratings overall?

- The RS-DMA met reliability criteria, as overall reliability of the RS-DMA was excellent across all items and raters (Fleiss Kappa= 0.82, 92.9% agreement, PABAK= 0.86)

How consistent are central raters in their ratings of each milestone?

- The RS-DMA met reliability criteria, with overall reliability of the RS-DMA was excellent across all items and raters (Fleiss Kappa= 0.82, 92.9% agreement, PABAK= 0.86)
- Substantial to excellent agreement among central raters for 27 of the 28 individual RS-DMA milestones, and a fair level of agreement for the milestone "Identifies body parts," due to challenges with feasibility when the eye-tracking overlay was not available (Table 1; Figure 3B)
- These findings indicate that milestone definitions are clear, reproducible, and applicable to the diverse functional presentations observed in RTT

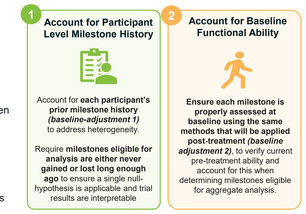
Validity: How comparable is the RS-DMA Historical Caregiver Interview vs. Performative RS-DMA for each milestone – Is Baseline Assessment Alone Enough to Control for Type 1 Error?

- There was a range of Slight to Excellent agreement between the RS-DMA Historical Caregiver Interview vs performative RS-DMA with an average of 81.8% and more than a third of items rating at Slight or Fair
- ***This discordance demonstrates that baseline assessment alone is not sufficient to control for Type 1 error*** *– discordance was due to greater reports of abilities from caregiver than were demonstrated during in-clinic assessments, highlighting the importance of controlling for both the historical report as well as performative assessment (for example at a trial baseline visit) to establish true milestone eligibility and control for type 1 errors* (Figure 4). This is particularly relevant in the context of a single-arm interventional trial in which the only control is each participant's pre-treatment milestone ability and expected spontaneous gain/regain based on natural history.
- Convergent validity between the RS-DMA and corresponding COAs items varied based on the similarity of how items were evaluated, with more similar items exhibiting higher agreement, confirming construct validity.

Clinical Meaningfulness

- 100% of the 28 milestones were endorsed by a majority of caregivers as independently meaningful as a potential treatment response, after observing the RS-DMA

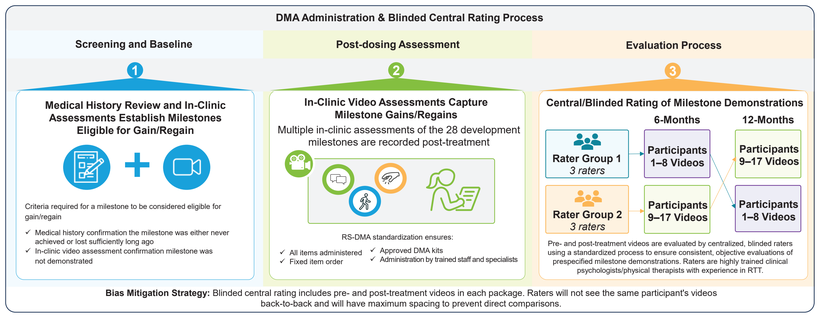
Figure 4: Controlling for type 1 errors



Conclusions

- The RS-DMA has strong psychometrics, that outperform all currently published clinical assessments for RTT, even in small sample size where even a single item or assessor difference can have dramatic impacts on psychometrics (Figure 2)
- The RS-DMA is validated for in-clinic assessment and video-based central rating which can be used in a single-arm design to mitigate bias from placebo effects, caregiver bias, at-home environments, and unstandardized administrations
  - Standardized in-clinic administration may result in type 2 errors (false negatives), which for truly disease modifying treatments is an acceptable potential signal loss in favor of controlling for type 1 errors (false positives)
- Using both the historical caregiver interview and pre-treatment performative assessment to establish a participant's baseline functioning is necessary to control for type 1 errors as our data shows that historical caregiver report's show higher milestone demonstration rates than performative assessments alone (Figure 4).
- Multiple stationary cameras and a wearable eye-tracker enabled scalable, feasible and reliable video-based central rating: certified raters can be deployed to maintain blinding across timepoints, rate the same videos to ensure validity and consistency, and mitigate recall bias by staggering participant videos, reinforcing the internal validity of a blinded, single-arm study design (Figure 5)
- Each milestone was clinically meaningful to caregivers, affirming the value using an individualized endpoint within the PABA framework to account for each participant's baseline function in considering if they are a responder to treatment
- When applied to the developmental plateau population, spontaneous gain and regain rates across the 28 milestones evaluated using the RS-DMA are sensitive to individualized change while remaining well-powered, ethical, and statistically rigorous in a small-sample single-arm design with a PABA analysis framework
- The RS-DMA offers a practical, bias-mitigated solution that can accelerate the evaluation of novel therapies while centering patient and caregiver perspectives, core tenets of value-based health outcomes research and FDA guidance on innovative trial designs in rare diseases (such as REVEAL)

Figure 5: DMA implementation process in the REVEAL pivotal trial



Key takeaways

- The RS-DMA is a psychometrically validated, trial-ready individualized endpoint for individuals with RTT that psychometrically outperforms all published RTT clinical outcomes assessments, even in a small-sample multiple-site validation study
- The RS-DMA's design and implementation methodologies ensure gold-standard psychometrics: strong treatment signal detection and noise reduction through tight control of inter-rater reliability, instrument construction, bias-mitigation, environment, and site- and patient-level variability
- Each milestone was rated as independently clinically meaningful
- To effectively control for false positives (Type 1 errors) within a small-sample, single-arm interventional study design using the plateau-anchored baseline-adjusted (PABA) analysis framework, milestone eligibility needs to account for both historical reported attainment and time-since loss of milestones and the performative demonstration of milestones in a bias-mitigated setting pre-treatment, to ensure a true treatment signal is detected post-treatment instead of false positives
- These findings provide psychometric validation of the RS-DMA in RTT for use as a regulatorily-aligned primary endpoint in innovative single-arm interventional study designs, such as the REVEAL pivotal trial
- The RS-DMA has been developed through multiple interactions with FDA who commented on "the thoughtful approach taken in designing the Developmental Milestone Assessment (DMA) tool" on its final presentation. In addition, the RS-DMA was appropriately tested as described here in the RESUME study to provide assurance the assessment tool is fit for purpose and is confirmatory Phase 3 trial ready

The authors wish to thank all patients and caregivers, as well as the study staff who make this research possible

Presenting author: Tessa Clarkson (tessaclarkson@psychlomere.com)

Acknowledgments: Study was funded by Taysha Gene Therapies. We would also like to acknowledge Julia Dailey, Jane Fiordelisio, Katie Vetro, Brendan Slagle, Rivka Cohen for their work supporting this project. Medical writing support for this poster was provided by Leigh O'Connor-Jones, of Avalere Health US Inc., funded by Taysha Gene Therapies
Disclosures: TC & MM are a paid consultants of Taysha Gene Therapies.

*Accessed from IRSF. ClinicalTrials.gov: NCT02738281: a prospective cohort of individuals with a pathologic mutation in the MECP2 gene, commonly associated with RTT. Cumulative incidence models of NHS data conducted by third-party statistical partners.

Abbreviations: CGI-S, clinical global impression– severity; COA: Clinical Outcome Assessment; FDA: Food and Drug Administration; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; MSELA: Mullen Scale of Early Learning Assessment; ORCA: Observer Reported Communication Assessment; PABAK: Prevalence-Adjusted and Bias-Adjusted Kappa.; PAOC: Patient-As-Own-Control; RESUME: Rett Syndrome Utility Methods Evaluation; R-MBA: Revised Motor Behavior Assessment Scale; RS-NHS: RTT and RTT-related Disorders Natural History Study; RS-DMA: Rett Syndrome Developmental Milestone Assessment; RSGMS: Rett Syndrome Gross Motor Scale; RSHFS: Rett Syndrome Hand Function Measure; RTT: Rett Syndrome; SD: Standard Deviation

References: 1. FDA. Innovative Designs for Clinical Trials of Cellular and Gene Therapy Products in Small Populations (2025). FDA-2025-D-3403. Accessed April 2026; 2. Li Y, Izem R. *Ann Transl Med*. 2022;10(18):1034; 3.Brim RL, Miller FG. *J Med Ethics*. 2013;39:703–707; 4. Neul JL, et al. *J Neurodev Disord*. 2026 Mar 12;18(1):20; 5.Downs J, et al. *Dev Med Child Neurol*. 2010;52(9):817–823; 6. Downs J, et al. *PLoS One*. 2016;11(1):e0147555; 7. Neul JL, et al. *J Child Neurol*. 2015;30(13):1743–1748; 8. Raspa M, et al. *Am J Intellect Dev Disabil*. 2020;125(6):493–509 9. Neul JL, et al. *J Neurodev Disord*. 2026;18(1):20

SCAN ME