Data validation and derivation plan

Related to RDM F02 Data validation and derivation plan

Study name (acronym):	A Study to Explore the Safety, Tolerability, Pharmacokinetic Profile, and Potential Efficacy of Guanabenz in Patients With Early-Childhood Onset Vanishing White Matter (VWM)
PI (or delegate):	M.S. van der Knaap, professor pediatric neurology, Amsterdam UMC
Date of completion:	30-07-2025
Version	1.0

Introduction

This documenet describes the data derivation and validation plan for the guanabenz trial in Vanishing White Matter (VWM). It describes the procedures used to generate analysis-ready datasets and the quality control measures implemented to ensure data accuracy, consistency, and integrity throughout the study. The plan supports compliance with the study protocol and provides guidance for handling data in cases where protocol requirements are not fully met.

For an overview of all Castor EDC univariate data validation, see:"[…]". All multivariate data validations, derived calculations and automations can be found here "[…]".

Objectives

- Implement robust procedures for the collection, validation, and verification of clinical trial data to maintain completeness, accuracy, and consistency
- Standardize the derivation of analysis-ready datasets to support transparent, traceable, and compliant statistical analysis
- Provide clear guidance for handling protocol deviations, ensuring decisions are well-documented and justified

Data validation plan

Valdation and quality control of collected data is applied continuously throughout the trial and includes built-in validations and automated calculations within the Castor Electronic Data Capture (EDC) system. Built-in edit checks within Castor help prevent entry errors in real-time, while manual source data verification ensures alignment between entered data and original source documents. Queries are generated for inconsistent or missing data and are tracked through resolution to maintain a clear audit trail. Regular data reviews and interdisciplinary meetings support timely identification and correction of discrepancies. The validation process also includes periodic audits by monitors or safety boards to confirm data integrity and protocol compliance. An overview of data validation procedures is provided in **Table 1**.

Table 1: Overview procedures data validation

Data validation	Specification	Explanation
Quality control of collected data	Built-in Castor checks	Automated checks to prevent data entry errors (see separate csv files, mentioned in the introduction).
	Data verification signatures	Confirms reviewed and accurate data.
	Signed forms for approval by PI	PI approval ensures protocol compliance.
	Manual Source Data verification	Confirms accuracy against source records.
	Query resolution tracking	Ensures resolution and audit trial
	Detection of missing data	Castor flags incomplete fields requiring attention or clarification. Note: Some visit data are marked as incompleted, while they are in fact completed or not applicable. These are checked and signed-off.
	Cross-check of date logic	Verifies temporal consistency (e.g., protocol date before screening and consent).
	Outlier detection in individual assessments	Identifies unusual or abrupt changes in scores that may indicate data issues.
	Adverse Event (AE) coding consistency	Ensures AE terminology aligns with standard naming and CTCAE grading.
	AE grading either via the Amsterdam UMC AE grading system or the CTCAE grading system	If either one defines the AE as normal, it will be considered normal.
	Standardization of concomitant medication entries	Confirms uniform naming conventions for medications across records.
	Check for dates that stay the same across visits	Some dates, like date of loss of walking, should be consistently filled in across visits.
	Check for dates that are in the future (reference 01-07-2025)	Since no data will be derived from 01-07-2025 onwards, there should be no dates after this time. **Note: dates that are explicitly set as missing in Castor, will show as large dates in the raw export (e.g. years 2995-2999, with 01-01 for month and day).**
Proof audits	6-Monthly audits by the monitor and DSMB	Independent review of data quality and safety.
Interdiscplinary trial meeting	M.S. van der Knaap	Weekly trial team meetings ensure clinical and operational oversight.
	R.J. Verbeek	
	E. van den Berg	

	M. M. C. Voermans	
Interdisciplinary Castor EDC VWM registry and GBZ trial data managagement meeting	M.S. van der Knaap	Biweekly data management meetings ensure data accuracy, resolve discrepancies and oversee data quality control.
	M. C. Postema	
	R.J. van Voorst	

Note. Collected data refers to data collected prospectively from the guanabenz trial database. VWM registry data refers to previously collected natural history data from the VWM registry, which will be used in this study as historical controls.

Collected and reused clinical outcome data are reviewed for internal consistency. Questionnaires containing similar items (e.g., walking ability) completed at the same timepoint are assessed for consistency. A standardized operating procedure (SOP) is used for completing the Health Utility Index scores to ensure a systematic review of the data. This SOP can be found at the following location: "[…]".

Data derivation plan

Variables in the reused dataset will be mapped and standardized to match the trial dataset format (e.g., units, coding, variable names). Only variables similar to those in the trial will be used. Derived variables (e.g. disease duration) will be calculated using the same logic as in the trial dataset. All derivation logic is documented in the "VWM1_GBZ_multivariate_validations.csv" file. A data dictionary with each derived variable includes metadata detailing its source variables and calculation methods. This data dictionary can be found in : "[…]". Version control will be applied to all datasets and derivation scripts using the Castor audit trial. Files will be stored with corresponding dates and R script versions, ensuring each version is clearly identifiable and properly linked. Missing values will be documented, and imputations will be pre-specified in the Statistical Analysis Plan (SAP). If critical variables are missing or inconsistently reported, the case may be excluded from specific analyses, with justification recorded. An overview of critical protocol deviations leading to missing data is provided in **Table 2.** Pre-specified decisions and documentation expectations are provided to ensure consistency and transparency.

Table 2: Protocol deviation handling plan

Data deviation	Specification	Decision
Deviations of eligibility criteria	Unwilling to travel to Amsterdam/follow the trial protocol	Exclude participant from study. Cannot meet the site-based requirements.
	Cannot guarantee adherence to treatment and study visitis due to family situation	Exclude participant from study. Risk of protocol non-compliance.
	Concurrent trial participation	Exclude participant from study. Risk of counfounding results or safety concerns.

	Allergies/hypertensitivity to guanabenz or other components of the formulations used in this study	Exclude participant from study. Safety risk.
Visits outside allowed time window (±28 days annual visits, ±14 days three months visits)	Annual visits (M0, M12, M24, M36, M48)	Allow if within acceptable risk window (≤ 14 days). Document as protocol deviation.
	Three months visits (M03, M06, M09, M15, M18, M21, M27, M30, M33, M39, M42, M45)	Allow minor deviations (≤7 days) with documentation.
Missed mandatory procedures	Blood spot cards	Reschedule if possible. If irrecoverable, document as missing data.
	Blood puncture	Reschedule. If not recovered, document and assess impact in PK/PD analysis.
	Lumbare puncture	Reschedule if possible within visit window. Document reason for missing procedure.
	MRI assessment	Reschedule if possible within visit window, document reason for missing procedure.
	Clinicial assessments	Reschedule as soon as possible. If not possible, ensure other safety/clinical date compensates.
	Patient diaries	Encourage continuation, replace with recall interview if possible, document compliance level
Unscheduled visits		Document clearly (reason and procedures done). Use the data if relevant to safety or efficacy outcome and analyse separately.

Decision data deviations 30-07-2025

During the trial, several protocol deviations occurred due to unforeseen circumstances. We have compiled an overview of these deviations, the corresponding decisions (including their rationale), and the impact each has on the analysis.

Table 3: Deviations and decision log

Data deviation	Specification	Decision

Deviations of eligibility criteria	One patient was unwilling to travel to Amsterdam/ follow the trial protocol	We excluded the participant from the study as it could not meet the site-based requirements. No measurements are taken into the analysis.
	One patient could not guarantee adherence to treatment and study visitis due to family situation	We exclude participant from study as this led to non-compliance. No measurements are taken into the analysis.
Visits outside allowed time window (±28 days annual visits, ±14 days three months visits)	One patient missed the annual visit within the allowed time window.	The visit was 22 days delayed because of a lost passport.
	One patient missed the MRI of the annual visit within the allowed time window	The annual visit was in time but the MRI had to be postponedby more than six weeks because of sickness; measurements are included in the analysis and recorded as a protocol deviation.
	Six patients were out of window for the 3 months safety visit in Amsterdam.	The three month visit was too early in one patient and later in 5 patients (range -5 days up to 15 days, with a median of 8 days). For one patient this was a miscount by the trial team. For 2 patients the reason was they had a sibling in the trial and the visit was scheduled on the same day for both (so 1 sibling was within the window and 1 out of the window). The other 3 patients were related to personal circumstances. measurements are included in the analysis and documented accordingly.
	During the whole duration of the trial n=28 video consultations in 17 different patients were out of window (range -11 up to 49 days, with a median of 12 days).	N=5 were related to the fact that a family has 2 children in the trial (the videoconsultations were scheduled for both children on the same day). N=2 consultations are a miscount by the trial team. N=8 consultations in 3 patients were due to social/personal circumstances. N=13 were out of window in 10 patients; those were related to personal circumstances of the family.
	For one patient the MRI of the end of study visit was performed after the official end date of the trial.	The end of study MRI was delayed by 2 days because of malfunctioning of the MRI scanner; measurements are

		included in the analysis and documented accordingly.
Missed mandatory procedures	A batch of the blood spot cards (n=25) were lost at the pharmacy	Rescheduling was not possible, and the data is unrecoverable. It is documented as missing, with minimal expected impact on the analysis, as ample data are available in the other 3-monthly measurements during the entire trial.
	A batch of serum blood samples (n=87) were lost at the pharmacy	Rescheduling was not possible, and although the data may be recoverable if the samples are found, it is currently documented as missing. PK/PD analysis remains feasible.
	Across all patients and visits, a total of 97 patient diaries were not filled out (n=105, 0% progress), or incompletely filled out (n=17, 1-50% progress). Out of the 105 patients diaries that were not filled out, N=25 diaries were missing of the same patient from baseline up to M3 visit, probably due to misunderstanding. After repeated instructions at the M3 visit diaries were filled out.	Not critical for primary/secondary outcomes; data recorded as missing without imputation. During the video consultation that took place every 3 months, missing information was retrieved.
	Clinical assessment: Two patients were not be able to perfom the GMFM-88 assessment due to medical circumstances of the patients.	Marked as missing; no imputation; available data will be used in longitudinal models with appropriate handling of missingness.
Unscheduled visits	There are 3 unscheduled vistis by 3 patients.	The reasons and procedures for these unscheduled visits are documented. Data from these visits are excluded from the analysis, as they are not relevant to safety or efficacy outcomes.
Scoring of clinical questionnaires	CFCS, GMFCS, HUI, MACS	At the moment of data lock, the scoring of these questionnaires starts with 0. However, according to the official scoring system, the scoring should start with 1. We will adapt this after data lock.