



Article

Measuring What Matters in Trial Operations: Development and Validation of the Clinical Trial Site Performance Measure

Mattia Bozzetti ¹, Alessio Lo Cascio ², Daniele Napolitano ^{3,*}, Nicoletta Orgiana ^{3,*}, Vincenzina Mora ³, Stefania Fiorini ⁴, Giorgia Petrucci ⁵, Francesca Resente ⁶, Irene Baroni ⁷, Rosario Caruso ^{7,8,†} and Monica Guberti ^{9,‡,§} on behalf of the Performance Working Group

¹ Department of Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy; mattia.bozzetti@asst-cremona.it or mattia.bozzetti@students.uniroma2.eu

² Direction of Health Professions, La Maddalena Cancer Center, 90146 Palermo, Italy; locascio.alessio@lamaddalenanet.it

³ CEMAD, Fondazione Policlinico Gemelli IRCCS, 00168 Rome, Italy; vincenzina.mora@policlinicogemelli.it

⁴ Department of Cardiovascular, Neural and Metabolic Sciences, Istituto Auxologico Italiano, IRCCS San Luca Hospital, 20149 Milan, Italy; s.fiorini@auxologico.it

⁵ Operative Research Unit of Orthopaedic and Trauma Surgery, Fondazione Policlinico Universitario Campus Bio-Medico, 00128 Rome, Italy; g.petrucci@policlinicocampus.it

⁶ Department of Oncohematology, Presidio Infantile Regina Margherita, Azienda Ospedaliero Universitaria Città della Salute e della Scienza di Torino, 10126 Turin, Italy; francesca.resente@unito.it

⁷ Health Professions Research and Development Unit, IRCCS Policlinico San Donato, 20097 San Donato Milanese, Italy; irene.baroni@grupposandonato.it (I.B.); rosario.caruso@unimi.it (R.C.)

⁸ Department of Biomedical Sciences for Health, University of Milan, 20141 Milan, Italy

⁹ Allied Health Professions Directorate, Istituto Ortopedico Rizzoli, 40136 Bologna, Italy; monica.guberti@ior.it

* Correspondence: daniele.napolitano@policlinicogemelli.it (D.N.); nicoletta.orgiana@policlinicogemelli.it (N.O.)

† These authors contributed equally to this work.

‡ Collaborators/Membership of Performance Working Group is provided in Appendix A.

Abstract

Background/Objectives: The execution of clinical trials is increasingly constrained by operational complexity, regulatory requirements, and variability in site performance. These challenges have direct implications for the reliability of trial outcomes. However, standardized methods to evaluate site-level performance remain underdeveloped. This study introduces the Clinical Trial Site Performance Measure (CT-SPM), a novel framework designed to systematically capture site-level operational quality and to provide a scalable short form for routine monitoring. **Methods:** We conducted a multicenter study across six Italian academic hospitals (January–June 2025). Candidate performance indicators were identified through a systematic review and expert consultation, followed by validation and reduction using advanced statistical approaches, including factor modeling, ROC curve analysis, and nonparametric scaling methods. The CT-SPM was assessed for structural validity, discriminative capacity, and feasibility for use in real-world settings. **Results:** From 126 potential indicators, 18 were retained and organized into four domains: Participant Retention and Consent, Data Completeness and Timeliness, Adverse Event Reporting, and Protocol Compliance. A bifactor model revealed two higher-order dimensions (participant-facing and data-facing performance), highlighting the multidimensional nature of site operations. A short form comprising four items demonstrated good scalability and sufficient accuracy to identify underperforming sites. **Conclusions:** The CT-SPM represents an innovative, evidence-based instrument for monitoring trial execution at the site level. By linking methodological rigor with real-world applicability, it offers a practical solution for benchmarking, resource allocation, and regulatory compliance. This approach contributes to advancing clinical research by providing a standardized, data-driven method to evaluate and improve performance across networks.



Academic Editors: Dorota Sys and Anna Scholz

Received: 2 September 2025

Revised: 23 September 2025

Accepted: 25 September 2025

Published: 26 September 2025

Citation: Bozzetti, M.; Lo Cascio, A.; Napolitano, D.; Orgiana, N.; Mora, V.; Fiorini, S.; Petrucci, G.; Resente, F.; Baroni, I.; Caruso, R.; et al. Measuring What Matters in Trial Operations: Development and Validation of the Clinical Trial Site Performance Measure. *J. Clin. Med.* **2025**, *14*, 6839. <https://doi.org/10.3390/jcm14196839>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: clinical trials; site performance; real-world data; advanced statistical methods; regulatory challenges; operational metrics; quality monitoring; clinical epidemiology; clinical trial site performance measure

1. Introduction

Clinical trials are central to therapeutic innovation and evidence-based medicine [1] but remain among the most complex and resource-intensive forms of research [2,3]. They face increasing complexity, rising costs, and growing expectations driven by regulatory demands, safety requirements, and pressures from multiple stakeholders [4,5]. From an epidemiological perspective, these challenges directly affect the validity, transparency, and reproducibility of trial findings, raising the need for methodological innovation in how trials are monitored and assessed.

Regulatory bodies have sought to address performance variability through Good Clinical Practice (GCP) guidelines and the promotion of Risk-Based Monitoring (RBM), which encourages adaptive oversight based on objective metrics [6]. However RBM implementation remains limited, largely due to the absence of validated and reproducible performance indicators [7,8]. Given their information and regulation-intensive nature, clinical trials could benefit substantially from digital tools such as electronic data capture (EDC) systems and RBM supported by advanced analytics, which have the potential to improve operational efficiency, data quality, regulatory compliance, and organizational preparedness [9]. Yet, uneven adoption of these technologies across sites and regions perpetuates performance disparities and risks exacerbating digital inequalities, with both managerial and ethical implications [10,11].

Within clinical epidemiology, this evolution highlights the principles of *Quality by Design* (QbD) that are strengthened by the ICH E6 (R3) GCP guidelines [12], which emphasize prospectively building quality into clinical trials rather than relying solely on retrospective monitoring. When combined with QbD, digital tools can serve as enablers of bias control, ensuring that trial conduct maintains internal validity and supports the reproducibility of results across multicentre settings.

Site performance is a decisive factor for trial success [1,13,14]. Problems such as poor recruitment, delayed visits, low data quality, or protocol deviations compromise both epidemiological validity and ethical accountability, often leading to trial delays or premature termination [15–17]. Recruitment challenges are a major contributor, with approximately 11% of cardiovascular trials on ClinicalTrials.gov ending prematurely due to poor accrual, resulting in significant ethical and economic consequences [18]. The absence of standardized, validated performance metrics exacerbates these issues, and around 30% of trials fail or remain unpublished [19], often due to preventable operational problems rather than scientific flaws [8] leading to wasted resources and eroding patient trust [20,21].

Current monitoring practices frequently rely on subjective or arbitrary indicators [22], making it difficult to identify underperforming sites, implement targeted improvement strategies, or compare performance across studies [23,24]. Additional barriers include the sustainability of digital solutions over time, institutional and governance constraints, legal and ethical risks, cybersecurity concerns, and an incomplete understanding of the contextual factors that enable or hinder digital transformation in healthcare [25,26]. Implementing a structured risk management approach could enhance quality and safety [27].

Furthermore, without a shared definition of “high performance” and standardized thresholds, assessments remain inconsistent, affecting site selection and monitoring [28]. Leveraging real-world data and advanced statistical approaches could support predictive models for more objective, scalable, and reproducible evaluations [29].

Within this context, the creation of a validated, scalable tool for the objective measurement of clinical trial site performance represents not only a managerial improvement but also a methodological advance in clinical epidemiology. Such a tool can reduce preventable failures, improve the efficiency of resource deployment, and uphold the ethical responsibility to ensure that research investments translate into meaningful and reproducible health outcomes.

2. Materials and Methods

2.1. Aims

This study aimed to develop and validate the Clinical Trial Site Performance Measure (CT-SPM) as a methodological innovation in clinical epidemiology. Specifically, the objectives were to:

- (i) Construct a psychometrically robust framework capturing core domains of site-level operational performance relevant to trial validity and reproducibility;
- (ii) Identify a parsimonious core indicator set through Mokken Scale Analysis to enable scalable and transparent monitoring;
- (iii) Establish evidence-based cut-offs to distinguish optimal from suboptimal site performance, thereby supporting bias control and risk-based monitoring; and;
- (iv) Design the instrument for seamless integration into a digital application, facilitating real-time benchmarking across multicentre trials and alignment with principles of QbD and modern clinical research governance.

2.2. Setting

A cross-sectional study (January–June 2025) was conducted in six Italian research centers (IRCCSs, university and public hospitals). A pre-specified protocol was published [30].

2.3. Inclusion Criteria

All the clinical trials at participating centers were eligible, aiming to maximize generalizability and minimize phase/status bias.

2.4. Instrument Development

The development of the CT-SPM followed a structured, multi-phase process designed to ensure both scientific rigor and practical relevance (Figure 1). Developed instruments are available in Table S1 and Metric selection in Table S3. The CT-SPM was designed as a two-tier instrument, comprising a short-form core set for universal application and a scalable full form that can be tailored to the monitoring needs of specific trials.

The retained metrics were structured as behaviorally anchored items suitable for self-assessment using a five-point Likert scale (1 = “Not frequent” to 5 = “Highly frequent”). To ensure accurate and comprehensive evaluation, the CT-SPM was completed jointly by all team members involved in each clinical trial. This consensus-based approach minimized individual bias and captured shared operational experience. Each item response reflected the frequency with which specific practices or events occurred during trial conduct.

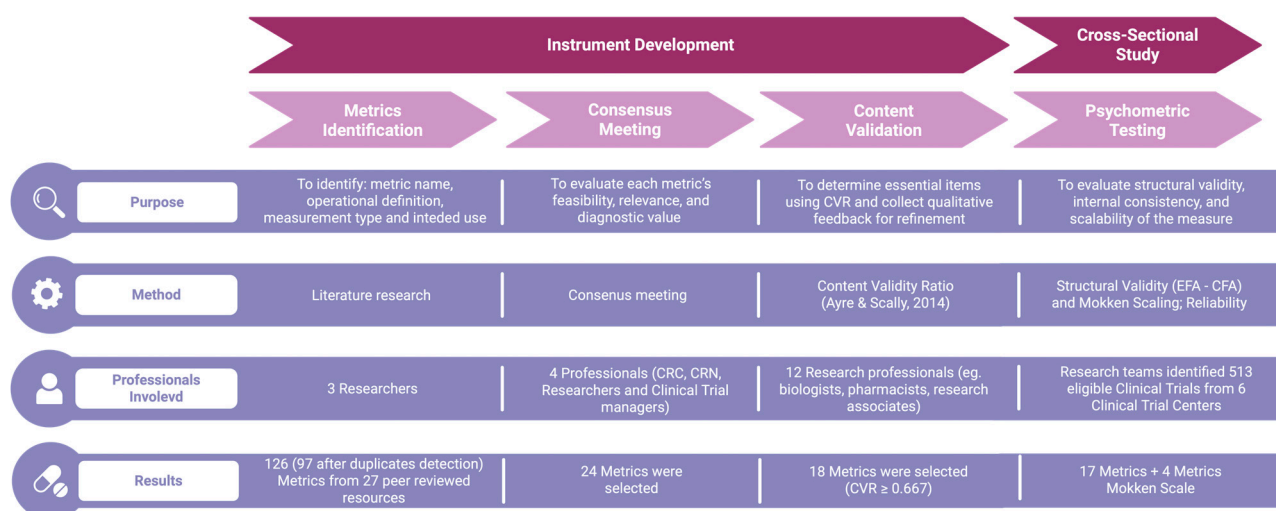


Figure 1. The figure illustrates the four sequential phases of the Clinical Trial Site Performance Measure (CT-SPM) development process: Metrics Identification, Consensus Meeting, Content Validation, and Psychometric Testing. For each phase, the diagram reports: the main purpose (e.g., identifying relevant metrics, evaluating content validity, assessing psychometric properties); the method used (e.g., literature review, expert panel discussions, Content Validity Ratio calculation, exploratory and confirmatory factor analysis); the types and number of professionals involved, and; the results obtained (e.g., number of metrics retained or refined). Abbreviations: CT-SPM = Clinical Trial Site Performance Measure; CRC = Clinical Research Coordinator; CRN = Clinical Research Nurse; CVR = Content Validity Ratio; EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis [31].

2.5. Variables Collected

A wide range of study-level and site-level variables were collected by the research team to avoid recall bias to characterize the context and to support exploratory analyses of associations with CT-SPM performance scores. Descriptive variables included the department or therapeutic area (e.g., oncology, haematology, cardiology), the type of treatment administered (single vs. multiple contact), trial phase (I–IV), translational nature (clinical vs. translational), and study design (e.g., randomized controlled trial, basket trial, platform trial, cluster trial). Information was also gathered on whether special procedures were applied during the study (e.g., GCP training, informed consent and documentation processes, adverse event reporting, complex data management procedures), and whether any centralized processes were implemented (e.g., centralized randomization, data management, or monitoring). In addition, a binary outcome variable was collected to serve as a reference criterion for performance classification. This variable was derived from a study-level question in which research teams were asked to indicate whether the trial, in their judgment, had deviated in a significant and unsatisfactory way from expected performance standards (1 = “Yes” and 0 = “No”).

2.6. Statistical Analyses

Data analysis proceeded in multiple stages. Descriptive statistics were first computed for each item, with results reported as means (M) and standard deviations (SD). Patterns of missing data were examined with Little’s missing completely at random (MCAR) test confirmed that data were MCAR [32]. To minimize potential loss of statistical power due to missingness, predictive mean matching (PMM) was used for imputation, implemented via the mice package [33]. Given the expected presence of ceiling and floor effects observed in the distribution of several items, a logarithmic transformation was applied to reduce

skewness and improve the approximation to normality. All analyses were conducted using R (version 4.3.3) [34] with statistical significance set at $p < 0.05$.

2.6.1. Psychometric Testing

Psychometric evaluation proceeded in multiple stages. Prior to exploratory factor analysis (EFA) analysis, data suitability was assessed using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy, with a minimum acceptable value set at 0.60, and Bartlett’s test of sphericity, with statistical significance established at $p < 0.05$. To cross-validate the factor structure, a split-sample approach was employed; the total sample was randomly divided into two subgroups, with approximately 60% allocated to EFA. Confirmatory Factor Analysis (CFA) was subsequently conducted on the second subsample to test the structural validity of a bifactor model composed of one general factor and two orthogonal specific factors. Model fit was evaluated using multiple indices: the chi-square test, Comparative Fit Index ($CFI \geq 0.95$), Tucker–Lewis Index ($TLI \geq 0.95$), Root Mean Square Error of Approximation ($RMSEA \leq 0.06$), and Standardized Root Mean Square Residual ($SRMR \leq 0.08$) [35]. The Bayesian Information Criterion (BIC) was also used for model comparison and to evaluate relative model parsimony.

Reliability was evaluated using several indices. McDonald’s omega (ω) and hierarchical omega (ω_h) were calculated to estimate internal consistency, with $\omega \geq 0.70$ indicating acceptable reliability. To assess unidimensionality, explained common variance (ECV) and ω_h values were examined, with thresholds of ≥ 0.70 used to support the use of a total score [36]. A lower ECV or ω_h suggested the need for scoring based on separate subscales.

2.6.2. Mokken Scaling

Mokken Scale Analysis (MSA) was conducted to develop a nonparametric short-form version of the CT-SPM. MSA comprises two main components: (a) an automated item selection procedure that organizes ordinal items into Mokken scales and (b) analytical techniques for testing the assumptions of nonparametric item response theory (IRT) models. Mokken models are based on three key assumptions: unidimensionality, local independence, and latent monotonicity (Ark, 2012 [37]; Mokken, 2011 [38]). Unidimensionality indicates that all items within a scale measure a single underlying construct (θ). Local independence implies that responses to one item are not influenced by responses to others. Latent monotonicity assumes that the probability of endorsing higher response categories increases monotonically with the latent trait.

The Monotone Homogeneity Model (MHM) incorporates all three assumptions and allows respondents to be meaningfully ranked along the latent dimension based on their total scores [39]. The Double Monotonicity Model (DMM) adds the requirement of invariant item ordering (IIO), whereby the rank order of item difficulty remains constant across all levels of the latent trait. Establishing such item hierarchies is essential to ensure generalizability and interpretability across populations.

To evaluate scalability, Loevinger’s scalability coefficients were computed. These include item-pair (H_{ij}), item (H_i), and scale-level (H_s) coefficients. Coefficients range from 0 to 1, with higher values indicating greater discriminative power. A scale is considered weak if H_s is between 0.30 and <0.40 , moderate if between 0.40 and <0.50 , and strong if ≥ 0.50 . Items with H_i values < 0.30 or those that violate IIO assumptions were considered for removal. The internal consistency of the scale was assessed using the Molenaar–Sijtsma reliability coefficient (ρ) [39]. Assumptions of monotonicity were tested using coefficient H and associated violation metrics. Monotonicity ensures that item scores increase proportionally with the underlying trait level, indicating consistent ordering along the latent continuum.

2.6.3. Calculation of Standardized Scores

For each metric and each factor, the standardized score was calculated using the formula:

$$\left(\frac{(\sum x - n)}{(n \times (5 - 1))} \right) \times 100$$

where x represents the score attributed to each individual item, while $\sum x$ denotes the total sum of item scores within a given factor. The variable n indicates the number of items that compose the factor. The constant 4 reflects the range of the Likert scale. The resulting value is then multiplied by 100 to express the standardized score on a 0–100 scale, facilitating direct comparisons across factors.

2.6.4. Performance Cut-Off Definition

Receiver Operating Characteristic (ROC) curve analysis was conducted to evaluate the discriminative capacity of the CTSPM and short form. The Area Under the Curve (AUC) was used as a summary index of diagnostic accuracy. Optimal cut-off values were determined using Youden's J statistic, identifying the threshold that maximized the trade-off between sensitivity and specificity in classifying trial protocols as adequate or inadequate in terms of site performance.

2.6.5. Sample Size Calculation

Sample size requirements for the MSA were determined using a power-based simulation approach to ensure sufficient statistical power to detect meaningful scalability under the assumptions of the Mokken model [40]. The simulation parameters included a baseline of 24 items, an initial sample size of $n = 200$ participants, and $n = 100$ Monte Carlo simulations per iteration. A target power of 0.80 was set, with the minimum acceptable scale scalability coefficient (Hs) fixed at 0.40. A maximum of 20 iterations was specified.

At each iteration, the average achieved power was calculated. If the desired power threshold was not met, the sample size was increased by 10 participants and simulations were repeated. This process continued until the power estimate reached at least 0.80 or the maximum number of iterations was exhausted. Results from these simulations indicated that a minimum sample size of 400 was required to achieve adequate power within 20 iterations, while increasing to 600 participants was necessary for simulations conducted over 40 iterations. These findings guided the target sample size for the MSA and ensured adequate model stability and generalizability of the short-form scale.

2.7. Deployment

A proof-of-concept digital dashboard was developed using Streamlit (Python 3.13) to demonstrate real-time computation of CT-SPM composite scores and workload indices. The prototype ingested structured trial, patient, and staffing data from local .csv sources (simulating EDC/CTMS API outputs), performed on-the-fly psychometric scoring, and rendered interactive dashboards for site-level performance monitoring, protocol complexity assessment, and resource allocation.

3. Results

The study included a total of 513 clinical trials. The majority of studies were conducted in oncology settings (46.0%), followed by gastroenterology (17.3%) and cardiology-pneumology (5.8%). Other specialties represented included allergology, neurology, rheumatology, and others, each accounting for less than 5% of the total. Regarding the type of treatment, 93.2% of the studies involved active interventions, while 6.4% were observational studies, either single or multiple contact surveys.

In terms of study phase, a large portion were Phase III (38.0%), followed by Phase II (21.5%) and Phase I (10.2%). Most studies did not involve translational research (52.3%), although 47.4% did. The most common study design was randomized controlled trials (RCTs), representing 43.1% of the total, followed by prospective cohort studies (29.6%) and observational non-interventional studies (5.7%). Vast majority of studies enrolled between 1 and 50 subjects (91.8%). As for study duration, most had a length of 1–3 years (86.9%), followed by 6 months–1 year (8.5%) and 0–6 months (3.0%). Finally, 90.4% of studies reported no protocol deviations, while 9.4% acknowledged the presence of deviations.

3.1. Phase 2: Psychometric Testing

Factorability of the data was supported by the KMO (overall MSA = 0.78; range 0.63–0.91 and a significant Bartlett's test of sphericity ($\chi^2(153) = 5970.97, p < 0.001$). Parallel analysis suggested retaining five factors. Subsequently, an EFA was performed using ML estimation with oblimin rotation to allow for correlated factors. The five-factor solution accounted for 60.3% of the total variance (Table S2).

However, during the CFA, the fifth factor and Item 11 showed unstable fit and did not meaningfully contribute to the model. Their inclusion negatively affected model parsimony and interpretability. Therefore, the fifth factor and Item 11 were excluded from the final CFA model, resulting in a more robust and theoretically coherent four-factor solution.

3.1.1. Structural Validity

The model demonstrated satisfactory fit to the data ($\chi^2_{(64)} = 1131.29, p < 0.001$; CFI = 0.983; TLI = 0.974; RMSEA = 0.079, 90% CI [0.071, 0.090]; SRMR = 0.066) accounting for 64.6% of the total variance. The measurement model comprises four first-order latent factors and two second-order latent factors. Factor 1 (F1), labelled “Participant Retention and Consent,” includes variables related to participant dropout, consent, and withdrawal rates. Factor 2 (F2), labelled “Data Completeness and Timeliness,” encompasses indicators of data quality, timeliness of Case Report Form entries, and rates of missing data. Factor 3 (F3), labelled “Adverse Events Reporting,” represents the frequency and accuracy of adverse event documentation. Factor 4 (F4), labelled “Protocol Compliance,” captures variables related to protocol deviations and violations.

At the higher level, the two second-order factors represent broader constructs. Factor G1, labelled “Participant Retention and Adverse Event Monitoring,” influences F1 and F3, reflecting participant-related outcomes and safety monitoring. Factor G2, labelled “Data Quality and Protocol Adherence,” influences F2 and F4, representing overarching data integrity and adherence to trial protocols (Figure 2). To improve model fit several residual covariances were specified based on high modification indices and theoretical plausibility. Specifically, residual correlations were added between Items 17–16 ($r = 0.42$), Items 17–4 ($r = 0.39$), Items 1–3 ($r = 0.34$), Items 6–4 ($r = 0.41$), and Items 4–13 ($r = 0.44$).

3.1.2. Reliability

Reliability analyses and the factor structure are summarized in Table 1.

Table 1. Reliability indices for the latent factors.

Factors	Omega Total (ω)	Omega Hierarchical (ω_h)	Explained Common Variance (ECV)
Participant Retention and Adverse Event Monitoring (G1)	0.73	0.59	0.42
Data Quality and Protocol Adherence (G2)	0.68	0.53	0.40
Participant Retention and Consent (F1)	0.82	0.70	
Data Completeness and Timeliness (F2)	0.75	0.64	
Adverse Events Reporting (F3)	0.70	0.60	
Protocol Compliance (F4)	0.65	0.58	

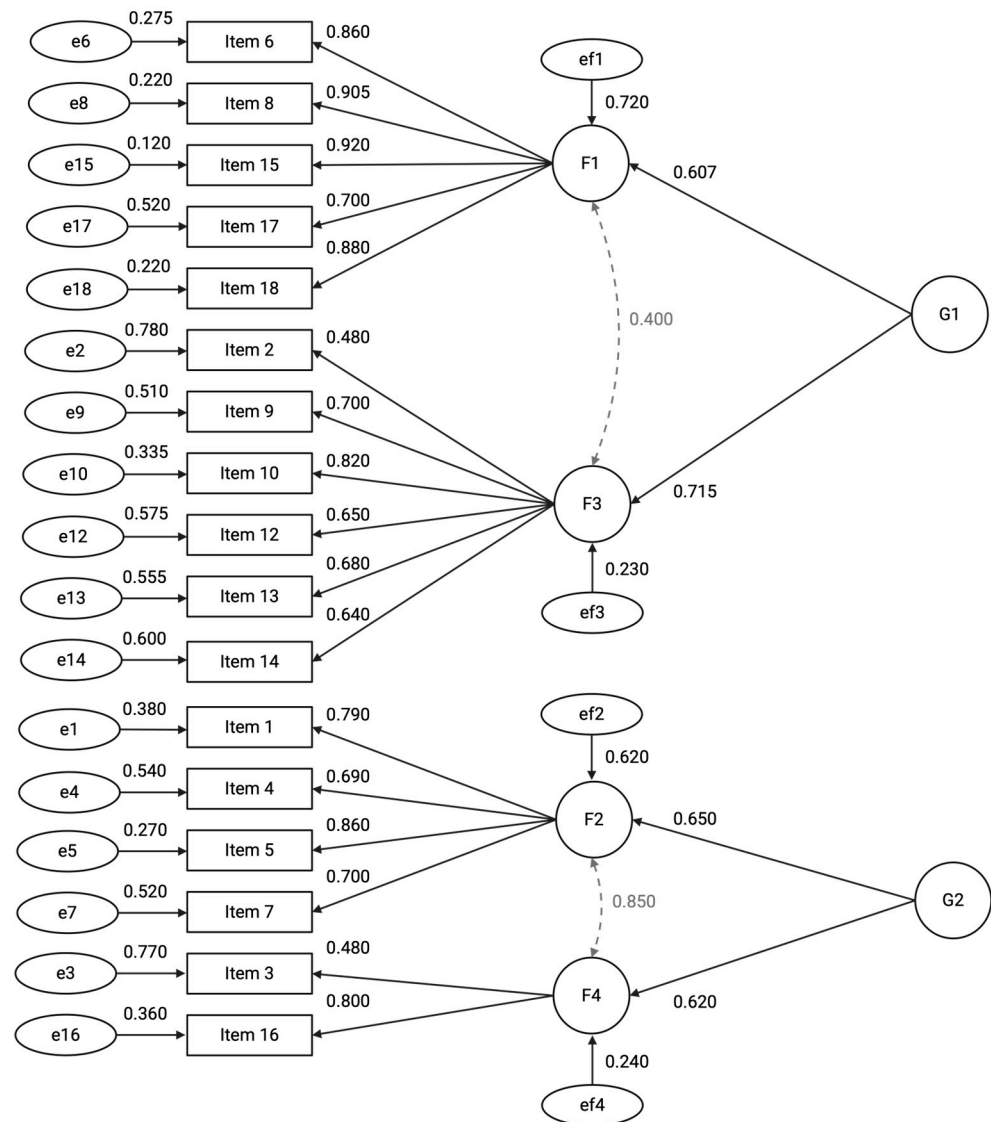


Figure 2. Path diagram of the hierarchical factor model illustrating relationships between latent constructs and observed variables. First-order factors (F1–F4) are indicated by circles, each measured by multiple observed items (rectangles) with standardized factor loadings shown on single-headed arrows. Second-order factors (G1 and G2) influence the first-order factors, with standardized regression coefficients labeled on the paths. Residual variances for observed items and first-order factors are represented by error terms (ellipses) with corresponding estimates. Bidirectional arrows denote covariances among first-order factors. All parameter estimates are fully standardized. For clarity, residual covariances between items—specified in the model—are not depicted in the diagram.

3.2. Phase 3: Mokken Scaling

3.2.1. Automatic Item Selection Procedure

The analysis resulted in the selection of a four-item scale composed of Item9, Item10, Item12, and Item17. The overall scalability coefficient of the scale was $H = 0.489$, $SE = 0.027$. H_i were as follows: Item9, $H_i = 0.563$; Item10, $H_i = 0.495$; Item12, $H_i = 0.557$; and Item17, $H_i = 0.504$. These values exceeded the conventional threshold of 0.30, supporting their inclusion. Item16, which showed $H_i = 0.216$, was excluded due to insufficient scalability.

3.2.2. Monotonicity

The item H_i were consistently above the acceptable threshold, ranging from $H_i = 0.548$ for Item17 to $H_i = 0.631$ for Item9. Minor violations of monotonicity were detected for

Item12 (1 violation) and Item17 (2 violations). No violations were reported for Item9 or Item10. Overall, the pattern of results suggests that the monotonicity assumption was largely met across the selected items.

Monotonicity was evaluated for the refined four-item Mokken scale composed of Item9, Item10, Item12, and Item17. Hi were consistently above the conventional threshold of 0.30, ranging from $H_i = 0.548$ for Item17 to $H_i = 0.631$ for Item9, indicating adequate scalability. Minor violations of monotonicity were detected for Item12 (1 violation) and Item17 (2 violations), while no violations were observed for Item9 or Item10. These findings suggest that the assumption of monotonicity was largely upheld across the selected items.

The results of the IIO analysis are presented in Table 2. The total HT coefficient was 0.461, reflecting a moderate to strong level of invariant ordering. Item-level HT values were uniformly 0.46, except for Item9, which exhibited a substantially stronger ordering structure ($HT = 0.82$). During the backward selection procedure, Item17 was flagged for removal due to two violations of invariant ordering. Nonetheless, given its acceptable scalability ($H_i = 0.53$) and theoretical relevance, it was retained in the final model. No other items demonstrated critical violations. Based on scalability, monotonicity, and invariant ordering, the final scale retained four items—Item9, Item10, Item12, and Item17.

Table 2. Summary of invariant item ordering (IIO) analysis for the final four-item Mokken scale.

Items	H	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	tmax	#tsig	crit	Selection	HT	Rho
Item9	0.63	4	1	0.25	0.18	0.18	0.05	1.14	0	1.96	0	0.55	0.82
Item10	0.53	5	0	0.00	0.00	0.00	0.00	0.00	0	1.96	0	0.60	
Item12	0.58	5	1	0.20	0.44	0.62	0.12	2.75	1	1.96	1	0.54	
Item17	0.53	4	2	0.50	0.57	0.70	0.18	5.05	1	1.96	2	0.63	

Note. H = item scalability coefficient, indicating the strength of the item's contribution to the hierarchical scale; #ac = number of active score groups; #vi = number of violations of invariant item ordering (IIO); #vi/#ac = proportion of violations relative to the number of active score groups; maxvi = maximum single violation; sum = total violations observed across score groups; sum/#ac = average number of violations per score group; tmax = maximum t-value observed for any violation; #tsig = number of statistically significant violations ($p < 0.05$); crit = critical t-value threshold at $\alpha = 0.05$; Selection = backward selection flag (1 = selected after backward elimination); HT = Loevinger's HT coefficient, representing the degree of invariant item ordering; Rho = Molenaar-Sijtsma reliability coefficient.

Internal consistency of the refined four-item scale was supported by $\rho = 0.818$, Cronbach's $\alpha = 0.778$, and Guttman's $\lambda_2 = 0.781$. All coefficients supported the psychometric adequacy of the refined scale (Table 3).

3.3. Phase 3: Defining Performance Cut-Off

The cut-off values varied across factors, reflecting differences in scoring distributions. The discriminative ability of the factors, measured by the AUC, ranged from poor to moderate, with the short form showing the highest AUC = 0.628. These results are summarized in Table 3.

Table 3. Cut-offs for each sub-scale.

Factors	M (SD)	Cutoff (Youden's)	AUC
Participant Retention and Consent (F1)	61.30 (25.70)	22.5	0.583
Data Completeness and Timeliness (F2)	71.80 (21.20)	46.9	0.528
Adverse Events Reporting (F3)	62.50 (10.90)	64.6	0.557
Protocol Compliance (F4)	82.50 (17.30)	6.25	0.293
Participant Retention and Adverse Event Monitoring (G1)	61.90 (14.90)	55.2	0.581
Data Quality and Protocol Adherence (G2)	77.10 (16.70)	60.9	0.562
Short Form (Mokken Scale)	40.27 (24.33)	59.38	0.628

Note. M = Mean; SD = Standard Deviation; AUC = Area Under the Curve.

4. Discussion

This study offers a methodological contribution to clinical epidemiology by introducing and validating an RBM instrument that addresses the persistent absence of validated, reproducible, and scalable indicators for site-level performance.

The CT-SPM was conceived as a two-tier instrument: (i) a scalable full form that can be tailored to study-specific monitoring needs and (ii) short-form core set for universal use across interventional and observational studies. Beyond operational management, the framework is conceived to strengthen internal validity, transparency, and reproducibility of clinical trials across multicentre settings, core priorities for contemporary clinical epidemiology [12]. By leveraging an integrated psychometric approach, this tool goes beyond isolated site metrics [5,22,41–43] to deliver a multifactorial framework aligned with current priorities in clinical epidemiology—such as bias control and robust reporting standards.

The CT-SPM revealed a coherent four-factor structure encompassing critical domains of trial execution. Factor 1 (Retention and Consent) reflects participant engagement and protocol adherence, anchored by indicators like consent rate. Factor 2 (Data Timeliness and Completeness) captures CRF delays and missing data, key metrics of data quality and workflow efficiency. Factor 3 (AEs) reflects both the accuracy and frequency of safety reporting, with item convergence on SAEs and data concordance supporting its validity. AE underreporting is often linked to procedural issues rather than clinical factors [44]. Factor 4 (Protocol Compliance) includes items on violations and deviations, revealing systemic issues in eligibility and coordination. While overlapping with participant behavior, these reflect operational weaknesses. Consistent with prior frameworks [45], protocol adherence is vital for trial integrity, with deviations known to impact outcomes significantly [46,47].

The emergence of two higher-order dimensions as Retention & Safety and Data Integrity & Protocol Adherence, provides a parsimonious lens for targeted quality interventions and pre-specified decision rules at the site level [48].

The MSA distills performance into four robust indicators: query frequency, SAE accuracy, outcome-data queries, protocol violations—balancing feasibility with measurement rigor. While the AUC indicates moderate discriminative accuracy, the short form enables transparent, continuous monitoring and rapid screening consistent with risk-proportionate oversight and pre-trial site selection [7,49,50]. Item-level behavior aligns with known quality levers: high query frequency is a sensitive proxy for workflow inefficiencies [51]; SAE concordance captures safety-reporting fidelity where under-reporting often reflects procedural rather than clinical causes [44,52]; primary-outcome queries track endpoint integrity [53] and protocol violations signal compliance risks that can degrade trial integrity [45–47].

Despite minor monotonicity (2) and IIO issues, it showed acceptable scalability ($Hi = 0.504$) and theoretical relevance, given the known link between violations and trial integrity [45]. Together, these four items define a coherent “operational reliability” construct, supported by moderate-to-strong scalability and few monotonicity breaches. Their grouping under the Data/Protocol Integrity factor aligns with modern trial quality frameworks [48], enabling comprehensive yet focused site assessment. This concise scale balances process and outcome metrics, favoring objectivity and feasibility. It supports risk-based monitoring and pre-trial site selection by flagging data quality and protocol compliance issues, providing actionable insights for sponsors [7,49,50,54].

The CT-SPM is designed for digital integration. Alignment with EDC/CTMS and hospital data flows enables near real-time analytics, supports pre-specified RBM triggers, and facilitates transparent audit trails. This positions the tool within a broader Quality-by-Design and good clinical practice ecosystem, moving beyond retrospective monitoring towards proactive bias control through design-embedded quality safeguards [12]. Looking

ahead, integration with real-world data sources and advanced computational methods can extend CT-SPM from descriptive monitoring to predictive and prescriptive decision support, for example, using supervised models to anticipate site under-performance and central statistical monitoring to detect anomalous patterns [29,55]. Such extensions should be accompanied by governance and fairness safeguards to mitigate automation bias and digital inequities [10,11,25].

Our performance framework has reporting and reproducibility implications. A minimal reporting set for site-performance methods and results including item definitions, scoring rules, thresholds, and decision logic—would enhance methodological transparency and comparability across trials and settings. Public availability of the instrument, scoring code, and a data dictionary would further support reusability and replication, addressing calls to reduce avoidable waste and improve the value of clinical research [20,52].

4.1. Strengths and Limitations

The CT-SPM provides a transparent, psychometrically validated and reproducible instrument that spans participant- and data-facing domains central to trial validity. The combination of a multidimensional full scale and a parsimonious short form enables scalable deployment for risk-based monitoring and Quality-by-Design applications across heterogeneous settings.

Some limitations must be acknowledged. First, analyses are cross-sectional, limiting causal interpretation of performance–outcome relationships. Second, cut-offs were calibrated against a subjective site-level criterion, introducing potential reference bias; future work should adopt objective outcomes and pre-registered decision rules. Third, the moderate AUC of the short form and the relatively lower discrimination of the protocol-compliance domain suggest that context-sensitive thresholds, control-chart approaches, or time-updated indicators may better capture dynamic risk. Fourth, the study was conducted in Italian sites; cross-national validation is needed to support generalizability.

4.2. Implications for Future Research

Although the CT-SPM provides a validated structure and a concise short form, accounting for 64.6% of the total variance, it remains clear that despite explaining a substantial proportion of the variance, the instrument does not yet capture the full complexity of what constitutes good performance across all trial contexts. Future investigations should therefore aim to validate the instrument internationally, prospectively testing it across different healthcare systems and therapeutic areas. The indicator set should be iteratively expanded and refined, with the development of context-specific metrics and the calibration of dynamic or control-chart-based thresholds that better reflect changes in site performance over time. Further research should also evaluate predictive and prescriptive approaches, such as supervised models and central statistical monitoring, to detect under-performance and anomalous patterns. In addition, the potential use of the CT-SPM as a design-embedded covariate (for example through stratification, adjustment, or hierarchical modelling with site-level random effects) deserves careful assessment. Finally, integration with RWD should be accompanied by explicit governance and fairness safeguards to mitigate automation bias and digital inequities. These steps will help move the CT-SPM from a proof-of-concept toward a next-generation performance score able to encompass the broader and more nuanced dimensions of trial quality.

5. Conclusions

The CT-SPM provides a structured, validated tool for assessing clinical trial site performance. Its dual format supports flexible application in diverse monitoring contexts. Further validation will strengthen its utility for evidence-based oversight and site management.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm14196839/s1>, Table S1: Clinical Trial Performance Metrics; Table S2: Exploratory Factor Analysis; Table S3: Clinical Trial Performance Metrics -Short Form (Mokken Scale).

Author Contributions: M.B.: Writing—original draft, Project administration, Data curation, Conceptualization, Methodology, and Formal analysis. D.N.: Writing—original draft, Investigation, Writing—review & editing, and Data curation. N.O.: Writing—original draft and Investigation. V.M.: Investigation, Data curation, and Project administration. S.F.: Investigation and Data curation. G.P.: Investigation and Data curation. F.R.: Investigation and Data curation. I.B.: Investigation and Data curation. A.L.C.: Investigation and Data curation. R.C.: Writing—review & editing, Project administration, Data curation, Conceptualization, and Methodology. M.G.: Writing—review & editing, Project administration, Conceptualization, and Methodology. Working Group: Investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Institutional Review Board Statement was waived for this study because no experimental interventions were tested.

Informed Consent Statement: Patient consent was waived for this study because no experimental interventions were tested.

Data Availability Statement: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: During the preparation of this manuscript/study, the author(s) used ChatGPT 4.0, OpenAI for the purposes of language refinement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Performance Working Group: Valentina Accinno, Caterina Fanali, Ester Udugbor, Marica Di Curcio, Eleonora Ribaudi, Stefania Colan-tuono, Francesca Garibaldi, Corinne Tavano, Valeria Altamura, Alessia Leonetti, Maria Anna Teberino, Cristina Graziani, Luciana Giannone, Roberta Savini, Giulia Wlderck, Daniele Privitera, Silvia Elettra Re-vere, Giulia Paglione, Giada De Angeli, Nathasha Samali Udugampolage, Miriam Angolani.

References

1. Wu, K.; Wu, E.; DAndrea, M.; Chitale, N.; Lim, M.; Dabrowski, M.; Kantor, K.; Rangi, H.; Liu, R.; Garmhausen, M.; et al. Machine Learning Prediction of Clinical Trial Operational Efficiency. *AAPS J.* **2022**, *24*, 57. [CrossRef]
2. Getz, K.A.; Campo, R.A. Trends in clinical trial design complexity. *Nat. Rev. Drug Discov.* **2017**, *16*, 307–308. [CrossRef]
3. Unger, J.M.; Cook, E.; Tai, E.; Bleyer, A. The Role of Clinical Trial Participation in Cancer Research: Barriers, Evidence, and Strategies. *Am. Soc. Clin. Oncol. Educ. Book Am. Soc. Clin. Oncol. Annu. Meet.* **2016**, *35*, 185–198. [CrossRef]
4. Rojek, A.M.; Horby, P.W. Modernising epidemic science: Enabling patient-centred research during epidemics. *BMC Med.* **2016**, *14*, 212. [CrossRef]
5. Walker, K.F.; Turzanski, J.; Whitham, D.; Montgomery, A.; Duley, L. Monitoring performance of sites within multicentre randomised trials: A systematic review of performance metrics. *Trials* **2018**, *19*, 562. [CrossRef] [PubMed]
6. European Medicines Agency. Guideline for Good Clinical Practice E6(R2). Published Online 1 December 2016. Available online: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-good-clinical-practice-e6r2-step-5_en.pdf (accessed on 16 February 2025).

7. Adams, A.; Adelfio, A.; Barnes, B.; Berlien, R.; Branco, D.; Coogan, A.; Garson, L.; Ramirez, N.; Stansbury, N.; Stewart, J.; et al. Risk-Based Monitoring in Clinical Trials: 2021 Update. *Ther. Innov. Regul. Sci.* **2023**, *57*, 529–537. [CrossRef] [PubMed]
8. Tudur Smith, C.; Williamson, P.; Jones, A.; Smyth, A.; Hewer, S.L.; Gamble, C. Risk-proportionate clinical trial monitoring: An example approach from a non-commercial trials unit. *Trials* **2014**, *15*, 127. [CrossRef] [PubMed]
9. Lee, M.; Kim, K.; Shin, Y.; Lee, Y.; Kim, T.-J. Advancements in Electronic Medical Records for Clinical Trials: Enhancing Data Management and Research Efficiency. *Cancers* **2025**, *17*, 1552. [CrossRef]
10. Sedano, R.; Solitano, V.; Vuyyuru, S.K.; Yuan, Y.; Hanžel, J.; Ma, C.; Nardone, O.M.; Jairath, V. Artificial intelligence to revolutionize IBD clinical trials: A comprehensive review. *Ther. Adv. Gastroenterol.* **2025**, *18*, 17562848251321915. [CrossRef]
11. Youssef, A.; Nichol, A.A.; Martinez-Martin, N.; Larson, D.B.; Abramoff, M.; Wolf, R.M.; Char, D. Ethical Considerations in the Design and Conduct of Clinical Trials of Artificial Intelligence. *JAMA Netw. Open* **2024**, *7*, e2432482. [CrossRef]
12. European Medicines Agency. Guideline for Good Clinical Practice E6(R3). Published Online 6 January 2025. Available online: https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf (accessed on 22 May 2025).
13. Bozzetti, M.; Soncini, S.; Bassi, M.C.; Guberti, M. Assessment of Nursing Workload and Complexity Associated with Oncology Clinical Trials: A Scoping Review. *Semin. Oncol. Nurs.* **2024**, *40*, 151711. [CrossRef]
14. Stensland, K.D.; Damschroder, L.J.; Sales, A.E.; Schott, A.F.; Skolarus, T.A. Envisioning clinical trials as complex interventions. *Cancer* **2022**, *128*, 3145–3151. [CrossRef] [PubMed]
15. Duley, L.; Antman, K.; Arena, J.; Avezum, A.; Blumenthal, M.; Bosch, J.; Chrolavicius, S.; Li, T.; Ounpuu, S.; Perez, A.C.; et al. Specific barriers to the conduct of randomized trials. *Clin. Trials* **2008**, *5*, 40–48. [CrossRef] [PubMed]
16. Durden, K.; Hurley, P.; Butler, D.L.; Farner, A.; Shriver, S.P.; Fleury, M.E. Provider motivations and barriers to cancer clinical trial screening, referral, and operations: Findings from a survey. *Cancer* **2024**, *130*, 68–76. [CrossRef]
17. Tew, M.; Catchpool, M.; Furler, J.; Rue, K.; Clarke, P.; Manski-Nankervis, J.-A.; Dalziel, K. Site-specific factors associated with clinical trial recruitment efficiency in general practice settings: A comparative descriptive analysis. *Trials* **2023**, *24*, 164. [CrossRef]
18. Baldi, I.; Lanera, C.; Berchialla, P.; Gregori, D. Early termination of cardiovascular trials as a consequence of poor accrual: Analysis of ClinicalTrials.gov 2006–2015. *BMJ Open* **2017**, *7*, e013482. [CrossRef]
19. Kasenda, B.; von Elm, E.; You, J.; Blümle, A.; Tomonaga, Y.; Saccilotto, R.; Amstutz, A.; Bengough, T.; Meerpohl, J.J.; Stegert, M.; et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA* **2014**, *311*, 1045–1051. [CrossRef]
20. Janiaud, P.; Hemkens, L.G.; Ioannidis, J.P.A. Challenges and Lessons Learned From COVID-19 Trials: Should We Be Doing Clinical Trials Differently? *Can. J. Cardiol.* **2021**, *37*, 1353–1364. [CrossRef]
21. Singh, G.; Wague, A.; Arora, A.; Rao, V.; Ward, D.; Barry, J. Discontinuation and nonpublication of clinical trials in orthopaedic oncology. *J. Orthop. Surg.* **2024**, *19*, 121. [CrossRef] [PubMed]
22. Sinha, I.P.; Smyth, R.L.; Williamson, P.R. Using the Delphi technique to determine which outcomes to measure in clinical trials: Recommendations for the future based on a systematic review of existing studies. *PLoS Med.* **2011**, *8*, e1000393. [CrossRef]
23. Klatte, K.; Subramaniam, S.; Benkert, P.; Schulz, A.; Ehrlich, K.; Rösler, A.; Deschodt, M.; Fabbro, T.; Pauli-Magnus, C.; Briel, M. Development of a risk-tailored approach and dashboard for efficient management and monitoring of investigator-initiated trials. *BMC Med. Res. Methodol.* **2023**, *23*, 84. [CrossRef]
24. Yorke-Edwards, V.; Diaz-Montana, C.; Murray, M.L.; Sydes, M.R.; Love, S.B. Monitoring metrics over time: Why clinical trialists need to systematically collect site performance metrics. *Res. Methods Med. Health Sci.* **2023**, *4*, 124–135. [CrossRef]
25. Hanisch, M.; Goldsby, C.M.; Fabian, N.E.; Oehmichen, J. Digital governance: A conceptual framework and research agenda. *J. Bus. Res.* **2023**, *162*, 113777. [CrossRef]
26. Raimo, N.; De Turi, I.; Albergo, F.; Vitolla, F. The drivers of the digital transformation in the healthcare industry: An empirical analysis in Italian hospitals. *Technovation* **2023**, *121*, 102558. [CrossRef]
27. Lee, H.; Lee, H.; Baik, J.; Kim, H.; Kim, R. Failure mode and effects analysis drastically reduced potential risks in clinical trial conduct. *Drug Des. Dev. Ther.* **2017**, *11*, 3035–3043. [CrossRef] [PubMed]
28. De Pretto-Lazarova, A.; Fuchs, C.; van Eeuwijk, P.; Burri, C. Defining clinical trial quality from the perspective of resource-limited settings: A qualitative study based on interviews with investigators, sponsors, and monitors conducting clinical trials in sub-Saharan Africa. *PLoS Negl. Trop. Dis.* **2022**, *16*, e0010121. [CrossRef]
29. Hampel, H.; Li, G.; Mielke, M.M.; Galvin, J.E.; Kivipelto, M.; Santarnecchi, E.; Babiloni, C.; Devanarayan, V.; Tkatch, R.; Hu, Y.; et al. The impact of real-world evidence in implementing and optimizing Alzheimer’s disease care. *Med* **2025**, *6*, 100695. [CrossRef]
30. Bozzetti, M.; Caruso, R.; Soncini, S.; Guberti, M. Development of the clinical trial site performance metrics instrument: A study protocol. *MethodsX* **2025**, *14*, 103165. [CrossRef] [PubMed]
31. Ayre, C.; Scally, A.J. Critical Values for Lawshe’s Content Validity Ratio: Revisiting the Original Methods of Calculation. *Meas. Eval. Couns. Dev.* **2014**, *47*, 79–86. [CrossRef]
32. Little, R.J.A. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *J. Am. Stat. Assoc.* **1988**, *83*, 1198–1202. [CrossRef]

33. Van Buuren, S.; Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **2011**, *45*, 1–67. [\[CrossRef\]](#)
34. R Core Team. R: A Language and Environment for Statistical Computing. Published Online 2023. Available online: <https://www.R-project.org/> (accessed on 12 March 2024).
35. Hu, L.; Bentler, P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* **1999**, *6*, 1–55. [\[CrossRef\]](#)
36. Rodriguez, A.; Reise, S.P.; Haviland, M.G. Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychol. Methods* **2016**, *21*, 137–150. [\[CrossRef\]](#)
37. Van der Ark, L.A. New Developments in Mokken Scale Analysis in R. *J. Stat. Softw.* **2012**, *48*, 1–27. [\[CrossRef\]](#)
38. Mokken, R.J. *A Theory and Procedure of Scale Analysis: With Applications in Political Research*; Walter de Gruyter: Berlin, Germany, 2011.
39. Sijtsma, K.; van der Ark, L.A. A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *Br. J. Math. Stat. Psychol.* **2016**, *70*, 137–158. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Straat, J.H.; van der Ark, L.A.; Sijtsma, K. Minimum sample size requirements for Mokken scale analysis. *Educ. Psychol. Meas.* **2014**, *74*, 809–822. [\[CrossRef\]](#)
41. Harman, N.L.; Bruce, I.A.; Kirkham, J.J.; Tierney, S.; Callery, P.; O'Brien, K.; Bennett, A.M.D.; Chorbachi, R.; Hall, P.N.; Harding-Bell, A.; et al. The Importance of Integration of Stakeholder Views in Core Outcome Set Development: Otitis Media with Effusion in Children with Cleft Palate. *PLoS ONE* **2015**, *10*, e0129514. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Whitham, D.; Turzanski, J.; Bradshaw, L.; Clarke, M.; Culliford, L.; Duley, L.; Shaw, L.; Skea, Z.; Treweek, S.P.; Walker, K.; et al. Development of a standardised set of metrics for monitoring site performance in multicentre randomised trials: A Delphi study. *Trials* **2018**, *19*, 557. [\[CrossRef\]](#)
43. Williamson, P.R.; Altman, D.G.; Bagley, H.; Barnes, K.L.; Blazeby, J.M.; Brookes, S.T.; Clarke, M.; Gargon, E.; Gorst, S.; Harman, N.; et al. The COMET Handbook: Version 1.0. *Trials* **2017**, *18* (Suppl. S3), 280. [\[CrossRef\]](#)
44. Phillips, R.; Hazell, L.; Sauzet, O.; Cornelius, V. Analysis and reporting of adverse events in randomised controlled trials: A review. *BMJ Open* **2019**, *9*, e024537. [\[CrossRef\]](#)
45. Meeker-O'Connell, A.; Glessner, C.; Behm, M.; Mulinde, J.; Roach, N.; Sweeney, F.; Tenaerts, P.; Landray, M.J. Enhancing clinical evidence by proactively building quality into clinical trials. *Clin. Trials Lond. Engl.* **2016**, *13*, 439–444. [\[CrossRef\]](#)
46. Jomy, J.; Sharma, R.; Lu, R.; Chen, D.; Ataalla, P.; Kaushal, S.; Liu, Z.A.; Ye, X.Y.; Fairchild, A.; Nichol, A.; et al. Clinical impact of radiotherapy quality assurance results in contemporary cancer trials: A systematic review and meta-analysis. *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.* **2025**, *207*, 110875. [\[CrossRef\]](#)
47. Ohri, N.; Shen, X.; Dicker, A.P.; Doyle, L.A.; Harrison, A.S.; Showalter, T.N. Radiotherapy Protocol Deviations and Clinical Outcomes: A Meta-analysis of Cooperative Group Clinical Trials. *JNCI J. Natl. Cancer Inst.* **2013**, *105*, 387–393. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Buse, J.B.; Austin, C.P.; Johnston, S.C.; Lewis-Hall, F.; March, A.N.; Shore, C.K.; Tenaerts, P.; Rutter, J.L. A framework for assessing clinical trial site readiness. *J. Clin. Transl. Sci.* **2023**, *7*, e151. [\[CrossRef\]](#)
49. Dombernowsky, T.; Haedersdal, M.; Lassen, U.; Thomsen, S.F. Criteria for site selection in industry-sponsored clinical trials: A survey among decision-makers in biopharmaceutical companies and clinical research organizations. *Trials* **2019**, *20*, 708. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Lamberti, M.J.; Wilkinson, M.; Harper, B.; Morgan, C.; Getz, K. Assessing Study Start-up Practices, Performance, and Perceptions Among Sponsors and Contract Research Organizations. *Ther. Innov. Regul. Sci.* **2018**, *52*, 572–578. [\[CrossRef\]](#)
51. Califf, R.M.; Zarin, D.A.; Kramer, J.M.; Sherman, R.E.; Aberle, L.H.; Tasneem, A. Characteristics of Clinical Trials Registered in ClinicalTrials.gov, 2007–2010. *JAMA* **2012**, *307*, 1838–1847. [\[CrossRef\]](#)
52. Ioannidis, J.P.A.; Greenland, S.; Hlatky, M.A.; Khoury, M.J.; Macleod, M.R.; Moher, D.; Schulz, K.F.; Tibshirani, R. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* **2014**, *383*, 166–175. [\[CrossRef\]](#)
53. Getz, K.A.; Stergiopoulos, S.; Marlborough, M.; Whitehill, J.; Curran, M.; Kaitin, K.I. Quantifying the magnitude and cost of collecting extraneous protocol data. *Am. J. Ther.* **2015**, *22*, 117–124. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Agrafiotis, D.K.; Lobanov, V.S.; Farnum, M.A.; Yang, E.; Ciervo, J.; Walega, M.; Baumgart, A.; Mackey, A.J. Risk-based Monitoring of Clinical Trials: An Integrative Approach. *Clin. Ther.* **2018**, *40*, 1204–1212. [\[CrossRef\]](#)
55. Pogue, J.M.; Devereaux, P.J.; Thorlund, K.; Yusuf, S. Central statistical monitoring: Detecting fraud in clinical trials. *Clin. Trials* **2013**, *10*, 225–235. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.