International Journal of Clinical Trials
Patel M et al. Int J Clin Trials. 2026 Feb;13(1):85-90
http://www.ijclinicaltrials.com
pISSN 2349-3240 | eISSN 2349-3259
DOI: https://dx.doi.org/10.18203/2349-3259.ijct20260051

Systematic Review

A systematic review of statistical approaches in clinical studies for atopic dermatitis treatment

Maheshvari Patel*, Nayan Patel, Rutuja Patil, Shambhavi Srivastava

Novo Bliss Research Private Limited 313, Silver Radiance-4, Gota, Ahmedabad, Gujarat, India

Received: 21 August 2025
Revised: 14 November 2025
Accepted: 18 November 2025

***Correspondence:**
Dr. Maheshvari Patel,
E-mail: maheshvari@novobliss.in

ABSTRACT

Biostatistics plays a pivotal role in clinical research by supporting accurate data analysis and interpretation, which are essential for drawing valid and evidence-based conclusions. In studies of atopic dermatitis (AD), a chronic, relapsing inflammatory skin condition, statistical methods are critical for evaluating clinical outcomes such as SCORAD scores, skin hydration, and patient-reported symptoms. As the number of trials investigating the efficacy and safety of skin care products for AD continues to grow, the quality and appropriateness of the statistical techniques used in these studies warrant close examination. This systematic review assessed statistical methods employed in clinical trials of skin care interventions for AD published between January 2020 and April 2025. Following PRISMA guidelines, a comprehensive search of PubMed was conducted, and 16 studies meeting the inclusion criteria were analysed in detail. A wide range of statistical methods was reported, including t-tests, ANOVA, ANCOVA, chi-square tests, and non-parametric alternatives such as the Wilcoxon signed-rank and Mann–Whitney U tests. Some studies also applied advanced techniques like mixed-effects models and ROC analysis. While most methods were generally appropriate for the study designs, frequent shortcomings were observed in the testing of assumptions and consistency of statistical reporting. These methodological gaps may limit the interpretability and reproducibility of trial outcomes. The findings highlight the need for improved statistical rigor and greater transparency in the analysis and reporting of clinical trials assessing skin care products for AD, ultimately to support more robust and reliable conclusions in dermatological research.

Keywords: Atopic dermatitis, Skin care products, Statistics, Parametric and non-parametric tests

INTRODUCTION

Biostatistics is a cornerstone of clinical research, providing the tools necessary for accurate data analysis, interpretation, and the generation of reliable, evidence-based conclusions. In dermatology and particularly in studies investigating atopic dermatitis (AD), a chronic inflammatory skin condition, statistical methods play a critical role in validating treatment outcomes of skin care products such as moisturizers and barrier creams.[1-3] Descriptive statistics are used to summarize key clinical variables,[4] while inferential techniques, including both parametric (e.g., t-tests, ANOVA), and non-parametric tests (e.g., Mann–Whitney U, Wilcoxon signed-rank).[5-10] assess treatment efficacy between groups. The selection of appropriate methods is often guided by assumption testing, such as the Shapiro-Wilk test for normality.[11] Moreover, confidence intervals, alongside p-values, offer deeper insights into the precision and clinical relevance of estimates. As clinical trials on skin care interventions for AD continue to grow in number, the methodological rigor and transparency of statistical approaches used in these studies are increasingly important. This systematic review critically evaluates the statistical methods employed in recent AD skin care trials, focusing on their appropriateness, correct application, and quality of

reporting, to identify methodological strengths and areas for improvement in future dermatological research.

METHODS

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The objective was to synthesize and analyse the statistical methodologies employed in skin care research studies with Atopic Dermatitis disease.

A comprehensive literature search was performed using the PubMed database. The search strategy combined Medical Subject Headings (MeSH) and relevant keywords related to atopic dermatitis and statistical methods. The keywords used included: "Atopic Dermatitis Statistical Analysis". The search was limited to articles published between January 2020 to April 23, 2025.

The initial search resulted in 201 articles, with no duplicate entries identified. These articles went through a two-step screening process: first, the titles and abstracts were screened, resulting in 25 articles being included; then, a full-text review was conducted on those 25 articles, with special attention given to whether the statistical methods were clearly outlined. Based on the established inclusion and exclusion criteria, 09 number of articles were eliminated. In the end, 16 studies were selected for the final review, and comprehensive data extraction was performed on all 16 studies.

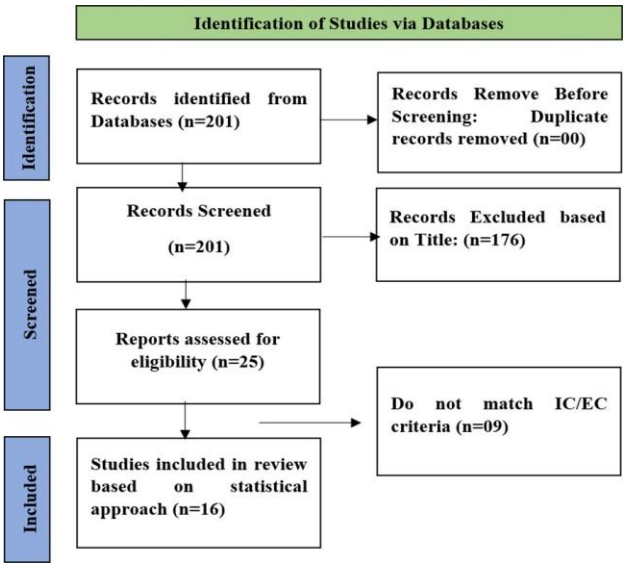

Figure 1: PRISMA flow diagram.

Inclusion criteria encompassed all full-text articles of prospective clinical trials with outcome data available involving human subjects in atopic dermatitis studies and description of statistical methods mentioned in articles. Exclusion criteria were non-human studies, pediatric population, non-English publications, without discussing statistical section and studies with incomplete or missing data, Unpublished or non-peer-reviewed data. Information collected included study characteristics (title, authors, publication year, journal, study design, sample size), patient characteristics (age & sex), intervention details (statistical methods), and outcomes (incidence, efficacy measures, safety). We systematically recorded results, including main findings. All data were systematically recorded, and findings were synthesized to identify trends and common statistical approaches in skin care research. The results were reported according to PRISMA guidelines, including a flow diagram outlining the study selection process and tables summarizing the key characteristics and findings of the included studies.

RESULTS

A variety of statistical methodologies were employed across the included 16 studies, tailored to data type, study design, and research objectives.

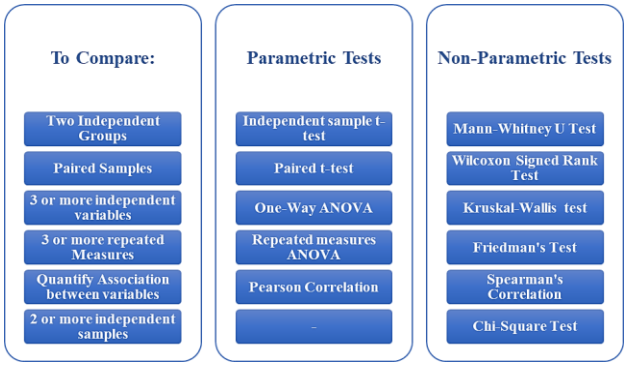

Figure 2: Classification of statistical procedure.

In a study of pH regulating Emollient cream on AD, data from 30 subjects were analysed to assess skin hydration, skin pH, and local SCORAD, with comparisons to baseline made using paired t-tests for each assessment time point and parameter.[5]

In a comparison study of AD vs alopecia areata for Deyo-Charlson Comorbidity Index score, categorical variables were presented as counts and percentages, with chi-square tests used to compare distributions between the AA and AD cohorts. For continuous variables like Clinical Burden parameters of AA Compared with AD were summarized using means, standard deviations (SDs), and medians, and comparisons between cohorts were made using two-tailed Student's t-tests.[4] In a randomized controlled trial, the Emollient group compare with control Emollient group, VAS pruritus score, patient global assessment score, SCORAD score, DLQI scores was compare by using Man-Whitney Rank Sum-test or two-independent t-test was used and for responder analyses (binary outcomes), a chi-squared or Fisher exact test was used for group comparison.[9]

A randomized study (n=60) comparing an emollient "plus" to 10% urea in AD patients, significant

improvements in TEWL and skin pH (weeks 4, 8, 12), and in SCORAD and hydration (weeks 8,12) favoured the emollient "plus" group. EASI, DLQI, and PVAS scores also differed significantly ae week 12. Analyses used repeated measure ANOVA and t-tests or nonparametric equivalents. Both treatments were well tolerated, with comparable baseline characteristics across.[6]

In a cross-sectional survey (n=401) of individuals ≥12 years, the χ^2 test with Yates' continuity correction and Fisher's exact test were used for the analysis of categorical variables in AD study included open questions, 5-point Likert-scale questions, multiple-choice questions, closed questions (yes/no), and evaluation scale questions (e.g., 1 to 10) and QoL has been investigated as feelings, stress and stigmatization.[12]

In an in vitro study, the cytotoxicity of FAT (125–2000 µg/ml) and ligustilide (0.5–50 µM) was assessed in HaCaT and RAW264.7 cells using MTT assays. Results (mean ± SD) were analysed using one-way ANOVA with Tukey's test.[7] A cross-sectional multicentre study conducted on 218 AD patients assessed by EQ-5D-3L, EQ-5D-5L, Dermatology Life Quality Index and Skindex-16, Investigator Global Assessment, EASI score and SCORAD score. Compare the difference in ceiling between 3L and 5L the McNemar's test was used also index score was tested by using Wilcoxon signed- rank test. A two-way random model with absolute agreement was used to estimate ICCs (interclass correlation coefficient).[10] The safety study comparing Upadacitinib versus Dupilumab treatments, Head and Neck Patient Global Impression of Severity (HN-PGIS) scale were analysed using the Cochran- Mantel-Haenszel (CMH) test, and EASI Scores were analysed using a mixed-effect model with repeated measures.[13]

In an open-label real-world study, the changes from baseline in xerosis, pruritus severity, sleep disturbance, and DLQI scores were analysed using an analysis of covariance (ANCOVA) model. ANCOVA was also applied for comparisons between treatment groups (cream vs. balm). Duration of itch was assessed using the log-rank test and Kaplan-Meier analysis.

The same statistical methods were used to perform ad hoc analysis of the change in xerosis and pruritus severity, duration of itch, sleep disturbance and DLQI score over the study period in two subpopulations.[14] For the emollient cream study on AD, 49 subjects were analysed using ANCOVA for parameters like skin barrier function (TEWL), skin moisturization (capacitance), skin surface dryness and EI (tolerability outcome), with treatment as a factor, subjects as a random effect, and covariates.[1] Similarly, for a moisturizer and topical Anti-Inflammatory drugs study with 33 AD subjects, multiple statistical tests, such as Dunnett's and Steel's multiple comparison tests, were used for analysis of VAS score, Sindex29 data and visual assessment score.[2]

A study on AD involving a cream containing a steroid and ceramide. skin barrier function, moisture levels, and severity were assessed. Within-group comparisons were analysed using the Mann-Whitney test and Scheffé's multiple comparison test. Between-group comparisons were evaluated using the Student's t-test, while the Holm- Bonferroni method was applied for multiple comparisons within groups.[3]

In a comparative analysis of redox homeostasis biomarkers in patients with psoriasis and AD. The distribution of variables in another study was assessed for normality using the Shapiro-Wilk test. Categorical variables were expressed as frequencies and percentages, while continuous variables were presented as medians with interquartile ranges (IQR). Group differences for categorical variables were analysed using the chi-square test. For continuous variables like biochemical parameters, the Mann-Whitney U test was used for comparisons between two groups, and the Kruskal-Wallis test was used for comparisons involving more than two groups. The diagnostic performance of the CAT in differentiating patients with psoriasis (PsO) from those with AD was evaluated using receiver operating characteristic (ROC) curve analysis.[11]

In a study evaluating a skincare product, parameters such as local SCORAD, itch intensity, skin hydration, and trans epidermal water loss (TEWL) were measured. The normality of the data was assessed using the D'Agostino & Pearson omnibus test. For non-parametric data, the Kruskal-Wallis test was used, followed by Dunn's multiple comparison test for post-hoc pairwise analysis. Data that met the assumption of normality but showed unequal variances were analysed using Welch's ANOVA, with Dunnett's T3 test used for subsequent pairwise comparisons. In vivo assessments included intra-group comparisons using the Wilcoxon signed-rank test and inter-group comparisons using the Mann-Whitney U test.[8]

In a study evaluating ruxolitinib cream for mild to moderate AD, key parameters such as symptom burden, overall quality of life (QoL), Patient-Oriented Eczema Measure (POEM), Dermatology Life Quality Index (DLQI), Children's DLQI (CDLQI), and Investigator's Global Assessment (IGA) were assessed. During the vehicle-controlled (VC) period, statistical significance was determined using analysis of covariance (ANCOVA). Categorical and responder analyses were conducted for POEM and EQ-5D-5L assessments, while post hoc analyses were performed for skin pain measured by the Numeric Rating Scale (NRS).[15]

In the real-world evaluation of AD patient, a total 118 patients enrolled during single visit the physical and patient-reported questionnaires are used for evaluations, this outcome data were presented and analysed by using mean and SD for normally distributed data and number and percentage for categorical data.[16]

Table 1: Statistical methods and key results.

S. no.	Statistical Test	Study Findings	References
01	T-tests	Skin hydration, skin pH, local SCORAD, clinical burden parameters, skin barrier function, moisture levels, severity	(5) (4) (3)
02	Mann-Whitney rank sum test or Wilcoxon signed rank test	VAS pruritus score, patient global assessment score, SCORAD score, DLQI scores, dermatology life quality index and skindex-16, investigator global assessment, EASI score	(9) (13) (11) (8) (10)
03	ANOVA	TEWL, skin hydration, EASI, DLQI, PVAS scores	(6) (7) (8)
04	ANCOVA	Xerosis, pruritus severity, sleep disturbance, DLQI scores, skin barrier function (TEWL), skin moisturization (capacitance), skin surface dryness, EI (tolerability outcome), QoL, POEM, DLQI, CDLQI, IGA	(14) (1) (15)
05	McNemar's test	EQ-5D-3L, EQ-5D-5L	(10)
06	Chi-square test	Deyo-charlson comorbidity index Score, questionnaires	(4) (12) (11)
07	Fisher's exact test	Open questions, 5-point likert-scale questions, multiple-choice questions, closed questions, evaluation scale questions, QoL	(12)
08	Shapiro-wilk test	To check normality	(11)
09	Kruskal-wallis test	To compare more than two groups	(11) (8)
10	Descriptive statistics or frequency and percentage	Physical and patient-reported questionnaires, categorical variables	(16) (11) (4)

DISCUSSION

This systematic review provides a comprehensive assessment of statistical methodologies employed in clinical and observational studies evaluating skin care interventions in AD. The findings reflect a broad application of statistical techniques ranging from basic descriptive measures to more advanced modeling approaches. While the majority of studies demonstrated a reasonable alignment between study design and analytical method, variability in the quality, transparency, and appropriateness of statistical application was frequently observed. Descriptive statistics formed the foundation of most studies, summarizing baseline characteristics, outcome measures, and adverse events. Continuous variables such as SCORAD scores, transepidermal water loss (TEWL), and skin hydration were typically presented using means and standard deviations or medians with interquartile ranges, depending on the data distribution.[4,6,11,16] Categorical variables, including treatment response and patient-reported outcomes, were commonly reported as frequencies and percentages.[4,12,13] For inferential statistics, both parametric and non-parametric methods were widely used, often appropriately selected based on data distribution and scale. Paired t-tests were employed for within-subject comparisons in longitudinal studies assessing changes from baseline, such as in the emollient cream trial evaluating local SCORAD and hydration levels.[5] Between-group comparisons were frequently conducted using two-tailed independent t- tests or the Mann–Whitney U test, depending on normality assumptions.[4,8,9] Notably, the use of assumption tests such as the Shapiro-Wilk test and D'Agostino & Pearson omnibus test was reported in some studies to validate the use of parametric tests.[8,11] indicating an awareness of statistical assumptions among researchers.

In studies involving repeated measurements over time, repeated measures ANOVA and mixed-effect models were utilized to account for intra-subject correlations and time effects.[6,13] These methods are particularly appropriate in randomized controlled trials with multiple assessment points, as they provide more efficient and unbiased estimates by incorporating both fixed and random effects. However, only a few studies explicitly reported model diagnostics or addressed potential violations of sphericity, which could compromise result validity if left untested. Analysis of covariance (ANCOVA) was also frequently used to adjust for baseline values and covariates, particularly in real-world or open-label studies comparing different product formulations.[1,14,15] This approach enhances statistical power and controls for confounding, but again, few studies described checks for interaction effects or the linearity of covariates with outcomes, important assumptions underlying ANCOVA models.

Non-parametric tests were correctly applied in several studies where data did not meet the assumptions of

normality or homogeneity of variance. For instance, Kruskal-Wallis tests followed by Dunn's post-hoc comparisons were used in studies involving multiple group comparisons with ordinal data.[8] The use of Wilcoxon signed-rank tests for intra-group comparisons and Mann-Whitney U tests for inter-group comparisons also appeared appropriately applied across several trials.[6,8,11] For categorical outcomes and responder analyses, chi-square tests and Fisher's exact tests were most commonly used.[9,12,13] In large samples, chi-square tests were preferred, while Fisher's exact test was applied in smaller samples where expected cell counts were low, ensuring accuracy of results. Some studies incorporated Yates' continuity correction to account for overestimation in chi-square tests with small sample sizes.[12]

Advanced statistical techniques such as ROC curve analysis were utilized to assess diagnostic performance in studies exploring biomarker differentiation between AD and psoriasis.[11] Similarly, intraclass correlation coefficients (ICCs) were calculated to assess agreement between scales like EQ-5D-3L and 5L, reflecting a thoughtful approach to evaluating reliability and measurement precision.[10] Post-hoc correction methods were used sparingly but appropriately, with Dunnett's, Holm- Bonferroni, and Scheffe's tests employed to control for type I error in multiple comparisons.[2,3,8] Several studies performed multiple hypothesis tests without applying any correction, raising concerns about inflated false-positive rates. The inconsistent reporting of confidence intervals, effect sizes, and power calculations across studies further limits the interpretability and clinical applicability of their findings.

Another key observation is the limited documentation of the rationale behind statistical method selection. While most studies employed appropriate techniques, few offered justifications. This lack of methodological transparency diminishes reproducibility and undermines confidence in the conclusions drawn. Furthermore, although real-world and survey-based studies often relied on descriptive statistics.[12,16] inferential testing was not always applied, potentially limiting the analytical depth and utility of the findings.

Finally, the use of more robust modeling approaches, such as generalized linear models or Bayesian methods, was notably absent from most studies. While not always necessary, such techniques could provide more nuanced insights, particularly in studies with complex designs, repeated measures, or multiple covariates. As the field moves toward precision dermatology, the adoption of more sophisticated and transparent statistical methodologies will be increasingly important.

CONCLUSION

This review shows that even though the statistical methods used in clinical studies of skin care products for atopic dermatitis (AD) were often suitable, there is still considerable need for improvement in transparency, assumption testing, and the clarity of reporting. Common techniques included parametric and non-parametric tests, t-tests, ANCOVA, repeated measures ANOVA, and mixed-effect models, which were applied with varying levels of detail and care. Yet, the limited use of assumption testing, lack of corrections for multiple comparisons, and insufficient reporting of effect sizes may reduce the reliability and interpretation of findings. To improve the strength and consistency of future research, studies should follow established statistical reporting standards, clearly explain their choice of methods, and apply techniques that match the study design and data type. Improving statistical practices will help raise the quality, trustworthiness, and clinical value of dermatological research.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Danby SG, Andrew PV, Taylor RN, Kay LJ, Chittock J, Pinnock A, et al. Different types of emollient cream exhibit diverse physiological effects on the skin barrier in adults with atopic dermatitis. Clin Exp Dermatol. 2022;47(6):1154-64.
2. Okoshi K, Ito S, Matsuoka M, Kinugasa Y, Shimizu E, Tanaka K, et al. Combination of a Topical Anti-Inflammatory Drug and a Moisturizer, Both with a Lamellar Structure Containing Synthetic Pseudo-Ceramides, for the Treatment of Patients with Mild-to-Moderate Atopic Dermatitis. Clin Cosmet Investig Dermatol. 2024;17:1569-78.
3. Matsuoka M, Okoshi K, Ito S, Kume T, Seki T, Nishizaka T, et al. Efficacy of Heparinoid Cream Containing Pseudo-Ceramide for Remission of Atopic Dermatitis. Clin Cosmet Investig Dermatol. 2021;14:1839-47.
4. Fenske DC, Ding Y, Morrow P, Smith SG, Silver MK, Moynihan M, et al. Comparing the burden of illness in patients with alopecia areata vs atopic dermatitis in the US population from a payer perspective. Manag Care Spec Pharm. 2023;29(4):409-19.
5. Ng SP, Bielfeldt S, Laing S, Danby S, Cork MJJ. Effects of a pH-Regulating Emollient Cream in Mild Atopic Dermatitis Patients with Moderate Localized Lesions. Skin Pharmacol Physiol. 2024;37(3):49-58.
6. Prakoeswa CRS, Huda BKN, Indrawati D, Umborowati MA, Anggraeni S, Damayanti, et al. Effectiveness and Tolerability of an Emollient "Plus" Compared to Urea 10% in Patients With Mild-to-Moderate Atopic Dermatitis. J Cosmet Dermatol. 2025;24(2):e70051.
7. Ha CW, Sohn EH, Kim SH, Jang S, Park MR, Kim YK, et al. Fermented Angelicae tenussimae with Aspergillus oryzae Improves Skin Barrier

Properties, Moisturizing, and Anti-Inflammatory Responses. Int J Mol Sci. 2022;23(20):12072.
8. Zhang Y, Heinemann N, Rademacher F, Darvin ME, Raab C, Keck CM, et al. Skin Care Product Rich in Antioxidants and Anti-Inflammatory Natural Compounds Reduces Itching and Inflammation in the Skin of Atopic Dermatitis Patients. Antioxidants. 2022;11(6):1071.
9. Magnolo N, Jaenicke T, Tsianakas A, Czech W, Thaçi D, Pinter A, et al. Comparison of different skin care regimens in patients with moderate to severe atopic dermatitis receiving systemic treatment: A randomized controlled trial. J Eur Acad Dermatol Venereol. 2023;37(5):18-26.
10. Koszorú K, Hajdu K, Brodszky V, Bató A, Gergely LH, Kovács A, et al. Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L descriptive systems and utilities in atopic dermatitis. Eur J Health Econ. 2023;24(1):139-52.
11. Klisic A, Bakic M, Karanikolic V. Comparative Analysis of Redox Homeostasis Biomarkers in Patients with Psoriasis and Atopic Dermatitis. Antioxidants. 2023;12(10):1875.
12. Stingeni L, Fortina BA, Baiardini I, Hansel K, Moretti D, Cipriani F. Atopic Dermatitis and Patient Perspectives: Insights of Bullying at School and Career Discrimination at Work. J Asthma Allergy. 2021;14:919-28.
13. Thyssen JP, Rosmarin D, Costanzo A, Warren R, Chu CY, Chovatiya R, et al. Efficacy and Safety of Upadacitinib versus Dupilumab Treatment for Moderate-to-Severe Atopic Dermatitis in Four Body Regions: Analysis from the Heads Up Study. Dermatology. 2025;241(1):10-8.
14. Deleuran M, Georgescu V, Decoster CJ. An Emollient Containing Aquaphilus dolomiae Extract is Effective in the Management of Xerosis and Pruritus: An International, Real-World Study. Dermatol Ther. 2020;10(5):1013-29.
15. Simpson EL, Augustin M, Thaçi D, Misery L, Armstrong AW, Blauvelt A, et al. Ruxolitinib Cream Monotherapy Improved Symptoms and Quality of Life in Adults and Adolescents with Mild- to-Moderate Atopic Dermatitis: Patient-Reported Outcomes from Two Phase III Studies. Am J Clin Dermatol. 2025;26(1):121-37.
16. Argenziano G, Mercuri SR, Savoia P, Amerio P, Fortina AB, Bongiorno MR, et al. Burden of Disease in the Real-Life Setting of Patients with Atopic Dermatitis: Italian Data From the MEASURE-AD Study. Dermatol Pract Concept. 2024;14(1): e2024079.

Cite this article as: Patel M, Patel N, Patil R, Srivastava S. A systematic review of statistical approaches in clinical studies for atopic dermatitis treatment. Int J Clin Trials 2026;13(1):85-90.