Allergy

WILEY

Allergy EUROPEAN JOURNAL OF ALLERGY AND CLINICAL IMMUNOLOGY EAACI

REVIEW

Statistical Analysis in Allergy and Immunology: A Review With Practical Examples

Michal Ordak[1] | Giovanni Paoletti[2,3] | Danilo Di Bona[4] | Bernardo Sousa-Pinto[5,6] | Matteo Martini[7,8] | Antonio Bognanni[2,3,9,10] | Jean Bousquet[11,12,13,14] | Giorgio Walter Canonica[2,3]

[1]Department of Pharmacotherapy and Pharmaceutical Care, Faculty of Pharmacy, Medical University of Warsaw, Warsaw, Poland | [2]Department of Biomedical Sciences, Humanitas University, Milan, Italy | [3]Personalized Medicine, Asthma and Allergy, IRCCS Humanitas Research Hospital, Milan, Italy | [4]Department of Medical and Surgical Sciences, School of Allergology and Clinical Immunology, University of Foggia, Foggia, Italy | [5]MEDCIDS - Department of Community Medicine, Information and Health Decision Sciences; Faculty of Medicine, University of Porto, Porto, Portugal | [6]CINTESIS@RISE – Health Research Network, Faculty of Medicine, University of Porto, Porto, Portugal | [7]Allergy Unit, Department of Internal Medicine, University Hospital AOU Delle Marche, Ancona, Italy | [8]Department of Clinical and Molecular Sciences, Marche Polytechnic University, Ancona, Italy | [9]Department of Health Research Methods, Evidence, and Impact & Department of Medicine, Evidence in Allergy Group, McMaster University, Hamilton, Ontario, Canada | [10]Clinical Epidemiology and Research Center (CERC), IRCCS Humanitas Research Hospital, Milan, Italy | [11]Institute of Allergology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität Zu Berlin, Berlin, Germany | [12]Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Immunology and Allergology, Berlin, Berlin, Germany | [13]Aria, Montpellier, France | [14]MASK-Air, Montpellier, France

Correspondence: Michal Ordak (michal.ordak@wum.edu.pl)

Received: 25 June 2025 | **Revised:** 21 July 2025 | **Accepted:** 6 August 2025

Funding: The authors received no specific funding for this work.

Keywords: allergy | immunology | statistical analysis

ABSTRACT

Statistical analysis plays a critical role in biomedical research, ensuring that data are interpreted appropriately and that conclusions are both valid and reproducible. In allergy and immunology, where studies increasingly rely on complex data structures and analytical approaches, clarity on biostatistical methods is essential to support transparency and scientific rigor. However, inconsistent statistical reporting and misuse of analytical techniques remain persistent challenges in the field. This review provides a structured and practice-oriented overview of key statistical aspects relevant to research in allergy and immunology Drawing upon recent peer-reviewed articles in these disciplines, we highlight best practices in the transparent reporting of statistical methods, verification of underlying assumptions, and interpretation of statistical significance in the context of clinical relevance. Each section is illustrated with practical examples to demonstrate sound analytical reasoning and to guide researchers, reviewers, and educators in improving statistical standards across the field.

Abbreviations: ACQ, Asthma Control Questionnaire; ANOVA, Analysis of variance; AQLQ, Asthma Quality of Life Questionnaire; CONSORT, Consolidated Standards of Reporting Trials; COVID-19, Coronavirus Disease 2019; DAGs, Directed Acyclic Graphs; FCS, Fully Conditional Specification; FeNO, Fractional exhaled Nitric Oxide; FEV_1, Forced Expiratory Volume in 1 seconds; GEE, Generalized Estimating Equations; GRADE, Grading of Recommendations Assessment, Development and Evaluation; IgE, Immunoglobulin E; LassoCV, Least Absolute Shrinkage and Selection Operator with Cross-Validation; MCID, Minimal Clinically Important Difference; MR, Mendelian Randomization; MR-Egger, Mendelian Randomization–Egger regression; MR-PRESSO, Mendelian Randomization Pleiotropy RESidual Sum and Outlier test; PCA, Principal component analysis; RNA-seq, RNA sequencing; ROUT, Robust Regression and Outlier Removal; SAMPL, Statistical Analyses and Methods in the Published Literature; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; VIFs, Variance Inflation Factors; WAME, World Association of Medical Editors; XGBoost, Extreme Gradient Boosting.

1 | Introduction

Biostatistics provides a fundamental framework for the planning, analyzing, and interpreting of biomedical data. It enables researchers to convert empirical observations into generalizable evidence and plays a critical role in informing decisions in both clinical and public health contexts. When applied appropriately, statistical methods allow for transparent reasoning and minimize the risk of misinterpretation. Conversely, poor statistical practices such as incorrect model selection, inadequate sample size calculation, selective or post hoc choice of statistical tests, inappropriate handling of missing data, overreliance on *p*-values without effect size interpretation, and failure to adhere to a predefined statistical analysis plan can lead to invalid conclusions and compromise the integrity of scientific findings [1, 2]. In addition, the growing complexity of data in allergy and immunology research highlights the importance of reproducible workflows, including transparent reporting of software and version, use of open-source tools when feasible, and sharing of analysis code to support reproducibility. Despite growing awareness of these issues, multiple studies have highlighted persistent weaknesses in statistical reporting across biomedical journals. A recent review demonstrated that the involvement of dedicated statistical reviewers remains limited, with no meaningful increase in adoption over the past two decades. Although statistical guidelines for authors and editorial boards have been published and widely disseminated, their consistent implementation in editorial workflows is lacking [3]. In a survey conducted among members of the World Association of Medical Editors (WAME), 40% of respondents estimated that only 31% to 50% of accepted manuscripts included statistically correct analyses. Notably, respondents with higher self-rated statistical expertise tended to be more critical in their evaluations. Moreover, the reported frequency of formal statistical peer review was low, with most editors estimating that only 1%–10% of submitted manuscripts undergo such specialized review. Frequently reported barriers included difficulty in identifying qualified reviewers and insufficient funding for statistical review [4]. These findings underscore the urgent need to improve statistical literacy, ensure the practical implementation of methodological guidelines, and strengthen the overall quality and transparency of statistical practices in biomedical research.

The aim of this review is to provide a structured and practice-oriented overview of statistical approaches commonly applied in allergy and immunology research. Drawing upon recent publications in the fields of allergy and immunology, we identify and discuss concepts that are particularly relevant to statistical practice in this research area. The review focuses on examples where statistical methods were applied in a clear, appropriate, and well-justified manner, demonstrating good practice in aligning the analytical approach with the research question and data structure. For each selected aspect, examples of good practice from the literature are presented along with commentary explaining the rationale behind the statistical decisions. Rather than offering a comprehensive guide to statistical analysis, this review aims to highlight key principles through practical examples from peer-reviewed research. Our goal is to support researchers, reviewers, and educators in enhancing transparency and quality of statistical applications across the field.

2 | Clear and Transparent Reporting of Statistical Methods

2.1 | Transparent Reporting of Statistical Methods: Rationale and Standards

Clear and transparent reporting of statistical methods is fundamental for the credibility, interpretability, and reproducibility of research in allergy and immunology. A well-documented statistical analysis allows readers to understand how conclusions were drawn, assess the validity of the findings, and replicate the results in future studies. Describing the analysis in alignment with the research question ensures methodological coherence, while precise definitions of variables and models enhance clarity. Furthermore, the appropriate use of statistical terminology such as distinguishing between "association" and "causation" or using "significance" in its technical sense prevents misinterpretation. Transparent reporting also supports peer reviewers and editors in evaluating the methodological rigor of a study. In clinical research, where findings may directly inform patient care or policy, such clarity is not only a matter of good practice but also of ethical responsibility. Ethical statistical practice includes avoiding selective reporting, p-hacking, and inappropriate data manipulation, all of which can undermine scientific integrity and public trust. Additionally, the growing use of advanced methods such as causal inference approaches beyond directed acyclic graphs (DAGs), machine learning applications for prediction, and statistical models that support personalized medicine highlights the need for continued methodological transparency and adaptation to evolving analytical challenges. However, it is important to emphasize that many of the most frequently observed statistical errors concern fundamental issues. These include insufficient evaluation of test assumptions, overinterpretation of *p*-values, incorrect or incomplete descriptive statistics, inadequate explanation of test purpose, and lack of adjustment for multiple comparisons. While advanced methods are often applied in collaboration with trained statisticians, these basic aspects of analysis remain the primary source of errors in many manuscripts [2, 3]. Several reporting guidelines and checklists, such as the CONSORT (Consolidated Standards of Reporting Trials), STROBE (Strengthening the Reporting of Observational Studies in Epidemiology), and SAMPL (Statistical Analyses and Methods in the Published Literature) guidelines, offer structured guidance for the transparent and comprehensive reporting of statistical analyses in biomedical research.

2.2 | Linking Statistical Methods to Study Design and Research Questions

The key elements of clear and transparent statistical reporting, along with illustrative examples from recent allergy and immunology studies, are summarized in Table 1. Each of these elements is further illustrated below with selected examples from the recent literature.

A good example of this level of transparency can be found in a study on asthma and long-term exposure to disinfectants and cleaning products, where the authors clearly described the use of advanced statistical methods such as latent class analysis and multiple imputation. They supported key analytical decisions

TABLE 1 | Key elements of clear and transparent statistical reporting with illustrative examples from recent allergy and immunology research.

Reporting element	Purpose and principle	Example from literature
Alignment of statistical analysis with research question	Ensures methodological coherence and interpretability	Latent class analysis and multiple imputation aligned with exposure assessment in asthma study [5]
Clear definition of statistical test purpose	Clarifies why a test is used and what it compares	Paired Wilcoxon test clearly linked to design in FPIES study [6]
Justification and documentation of variable selection	Enhances transparency and replicability of modeling decisions	Stepwise model with justified variables based on prior evidence in cow's milk immunotherapy study [7]
Use of appropriate descriptive statistics	Supports correct interpretation of distributions	Continuous variables expressed as medians and IQRs or means and 95% CI; categorical variables as counts and percentages, as reported in the FPIES study [6]
Reporting software and versions	Facilitates reproducibility and transparency	R v4.3.2, Scanpy v1.10.3, and GraphPad Prism 9 in eosinophilic asthma omics study [8]
Transparent handling of missing data	Allows readers to assess bias and modeling strategy	Missing indicator method used with rationale in post-COVID allergic disease cohort [9]
Reference to methodological sources and tools	Increases credibility of complex methods	Disinfectant exposure study cited software and methodology for imputation and latent class modeling [5]
Proper use of statistical terminology	Prevents misinterpretation and misuse of core statistical language	In the disinfectant exposure study, authors cautiously described effects as "in association with" asthma outcomes [5]

with references to methodological literature and software documentation, enhancing both the interpretability and reproducibility of their analysis. The model selection process, evaluation of model identification, and pooling of estimates across imputed datasets were reported in a structured and detailed manner. Such referencing helps guide readers through complex procedures and ensures that the applied methods are based on established statistical standards [5].

The purpose of a statistical test should be clearly described, as exemplified in a longitudinal pediatric study on food protein-induced enterocolitis syndrome (FPIES), where the authors specified not only the use of the paired Wilcoxon test but also its application to compare values between matched groups (e.g., allergic patients vs. controls) and within the same individuals over time (e.g., before vs. after oral food challenge). This level of detail ensures that the statistical method is directly linked to the study design and research question, enhancing both clarity and interpretability [6].

Another important aspect of statistical clarity and methodological transparency is the justification for how variables are selected and defined. This is well illustrated in a recent study on oral immunotherapy for cow's milk allergy in young children, where the authors explained that the variables included in the analysis were chosen based on previously published evidence and clinical relevance. The variables were described in terms of their timing, source, and clinical role, with each model built step by step and fully documented. Such an approach reinforces the transparency of the analysis and allows readers to assess the rationale behind key analytic decisions. It is therefore essential that researchers report not only which variables were used, but also why and how they were selected [7].

Transparent reporting should also include appropriate descriptive statistics that summarize the distribution and central tendency of variables, tailored to their scale and distribution. For example, using medians and interquartile ranges for skewed

continuous data, or frequencies and proportions for categorical variables, enhances the clarity and interpretability of study populations and outcomes. It is important to remember that clearly reporting the software used for statistical analysis, including version numbers, is a key aspect of methodological transparency. For example, in a study on eosinophilic inflammation in severe asthma, the authors specified that *R* (v4.3.2) was used for protein omics analysis, Scanpy (v1.10.3) in Python for single-cell RNA-seq data, and GraphPad Prism 9 for the analysis and visualization of experimental data. This level of detail facilitates reproducibility and increases confidence in the robustness of the findings [8].

2.3 | Addressing Data Limitations and Enhancing Reproducibility

An important aspect of statistical transparency is how missing data are handled and reported. For example, in a recent multinational cohort study on incident allergic diseases in post-COVID-19 condition, the authors explained that they used the missing indicator method, generating separate variables to account for missingness and including them in the adjustment set. This approach, supported by methodological references, helps reduce potential bias and allows readers to understand how missing data were incorporated into the analysis [9].

Another important aspect of statistical clarity and methodological transparency is the verification and fulfillment of assumptions underlying statistical tests and models. Ensuring that these assumptions are met is essential for the validity of the results and the correct interpretation of findings. As this is a more in-depth and technical issue, it is addressed separately in the following section.

3 | Verifying and Addressing Assumptions in Statistical Tests and Models

Verifying the assumptions underlying statistical tests is essential to ensure the validity of the results. When assumptions are violated and not addressed, estimates may become biased, confidence intervals misleading, and *p*-values unreliable. This can ultimately lead to incorrect conclusions, even when standard statistical methods are used. In clinical and epidemiological research, where findings may inform diagnosis, treatment, or public health decisions, such errors can have significant consequences. Despite this, assumptions and the process for assessing whether they are satisfied are often overlooked or unreported in published studies, which undermines the transparency and credibility of statistical analyses.

3.1 | Assumptions in Commonly Used Statistical Tests

Commonly used statistical tests, such as the t-test or ANOVA, rely on assumptions including normality of distribution, homogeneity of variances, and independence of observations. Assessing these assumptions helps ensure the validity of results, especially when sample sizes are not large enough for asymptotic approximations to apply with confidence. Key assumptions of commonly applied statistical tests are summarized in Table 2.

Equally important is the appropriate selection of post hoc tests, as the choice (e.g., Bonferroni, Tukey, or Games–Howell) should align with the specific assumptions about group variances and sample sizes to ensure valid pairwise comparisons. Bonferroni correction is particularly suitable when controlling the family-wise error rate across a limited number of planned comparisons, but may be overly conservative in settings with many tests, potentially inflating type II error. In such cases, alternative methods like Holm or Hochberg procedures may provide a better balance between error control and power. For example, in a randomized trial on sublingual immunotherapy in adolescents with allergic rhinitis, the authors used the Shapiro–Wilk and Levene tests to assess distributional normality and variance homogeneity in a sample of 65 participants. Depending on the results of these tests, they appropriately selected between the standard Student's t-test, Welch's t-test for unequal variances, or nonparametric alternatives. In the case of repeated measures, the authors applied repeated measures ANOVA, including preliminary tests of sphericity, and used Greenhouse–Geisser or Huynh–Feldt corrections when the assumption of sphericity was violated. This reflects a cautious and transparent approach to model assumptions in a moderately sized study and demonstrates careful matching of statistical techniques to the properties of the data [10]. When assumptions are uncertain or violated, robust methods like Welch's t-test or permutation-based approaches can provide more reliable results and help prevent misleading conclusions. Clear reporting of how assumptions are assessed and how methods are chosen accordingly enhances the transparency and reproducibility of statistical analysis.

3.2 | Interpreting Assumptions in Large Samples and the Role of Outliers

In studies with large sample sizes, it is a common misconception that parametric tests such as the t-test or ANOVA require the data to be normally distributed. In practice, with sufficiently large numbers of observations, moderate departures from normality have little influence on the validity of the results. In these situations, it is important to consider whether the variances between groups are similar and whether the data points are independent. These properties are determined by the overall structure and design of the study. If variances differ meaningfully, alternative versions of standard tests that do not assume equal variances can be applied to maintain the validity and robustness of statistical inferences. As a result, preliminary tests for normality in large samples often add little value and can unnecessarily complicate the analysis [11].

It is important to remember the role of outliers, which can significantly distort the results of statistical tests based on means. A clear example is provided by a study on the long-term efficacy of cow's milk anaphylaxis immunotherapy, in which the authors explicitly reported that substantial departures from normality and from the assumption of no outliers were present [12]. In such cases, where outliers and non-normal distributions are present, it is important to consider how these characteristics may influence statistical estimates and test selection. Evaluating

TABLE 2 | Assumptions in commonly used statistical tests: Key principles and examples from allergy and immunology research.

Assumption-related element	Purpose and principle	Example from literature
Assessment of normality and variance homogeneity	Ensures validity of parametric tests by verifying key assumptions	Sublingual immunotherapy study used Shapiro–Wilk and Levene tests [10]
Appropriate test selection based on assumption checks	Improves inference accuracy by aligning method with data properties	Same study used Welch's t-test or nonparametric alternatives as needed [10]
Corrections for sphericity in repeated measures ANOVA	Maintains validity when sphericity is violated	Greenhouse–Geisser and Huynh–Feldt corrections used in repeated measures design [10]
Limited role of normality tests in large samples	Avoids unnecessary tests; focuses on relevant assumptions	Large sample analysis: attention focused on variance comparability and independence [11]
Identification and reporting of outliers	Enhances transparency and robustness	Cow's milk anaphylaxis immunotherapy study explicitly reported substantial departures from normality and from the assumption of no outliers [12]
Matching method to measurement scale	Ensures correct methodological decision and interpretability	Hair dye allergy study used Wilcoxon signed-rank test for ordinal data [13]
Method adaptation to small sample sizes and non-normal data	Enables valid analysis despite constraints	Nonparametric methods (Mann–Whitney U, Spearman) used in allergy–nutrition study with small sample size [14]

the robustness of results by comparing alternative methods or summary measures can help ensure that the conclusions remain valid [1]. While removing outliers may sometimes be justified, such as when they result from measurement error, it should be done with caution. In many cases, robust statistical methods (e.g., trimmed means, M-estimators, or nonparametric tests) allow for meaningful analysis without excluding data points that may represent true variability. Clear justification and transparent reporting are essential when addressing outliers.

An example of appropriate consideration of measurement scale is a clinical study investigating a permanent hair dye designed to reduce allergic contact dermatitis, where ordinal data provided by investigators, cosmetologists, and participants were analyzed using the Wilcoxon signed-rank test. This reflects a correct methodological decision that aligns the statistical approach with the level of measurement [13].

3.3 | Matching Test Selection to Sample Size, Data Type, and Distribution

Ensuring that test selection matches the nature of the data is essential for drawing valid and interpretable conclusions. Sample size and the balance between comparison groups in terms of baseline characteristics are important considerations when choosing statistical methods. For instance, in a study investigating the impact of dietary fat intake and breast milk composition on allergy outcomes in offspring, the authors explicitly stated that due to small sample sizes and generally non-normal distributions, nonparametric tests were applied. Specifically, the Mann–Whitney U test was used for group comparisons, and Spearman's rank correlation was used to assess associations, reflecting an appropriate adaptation of methods to data constraints [14].

3.4 | Methodological Considerations in Advanced Regression and Causal Modeling

Assumptions play a critical role not only in simple statistical tests but also in regression models and more advanced analytical approaches. Each type of model is associated with its own set of assumptions, which must be checked and addressed to ensure valid inference. Violations of key assumptions such as linearity, independence of errors, or absence of multicollinearity can lead to biased estimates, incorrect standard errors, and misleading conclusions. As regression-based and complex methods are widely used in allergy and immunology research, careful evaluation and transparent reporting of whether their underlying assumptions are met is essential for scientific validity. Key assumptions and reporting practices relevant to regression and advanced statistical methods are summarized in Table 3.

One important consideration when using advanced statistical methods is the approach to variable selection, which itself involves specific assumptions. It is essential to ensure that the chosen selection method aligns with the study objective and the underlying data structure. For example, in a study on maternal dietary avoidance in infants with food protein-induced allergic proctocolitis, the authors used several methods

TABLE 3 | Key assumptions and reporting practices in regression and advanced statistical methods with examples from allergy and immunology research.

Assumption-related element	Purpose and principle	Example from literature
Transparent approach to variable selection	Aligns method with study goals and data structure	Proctocolitis study used Lasso, logistic regression, XGBoost, Random Forest for variable ranking [15]
Accounting for non-linearity and interactions	Captures complex associations and avoids residual confounding	Oxidative balance and asthma study used restricted cubic splines and subgroup analyses [16]
Avoiding arbitrary categorization of continuous variables	Preserves variability and statistical precision	Methodological article discourages categorizing continuous variables and promotes use of splines [17]
Use of causal diagrams and structured confounding control	Supports valid inference through prespecified causal logic	DAG and overlap weighting in asthma-biologics study [18]
Use of robust standard errors in regression	Improves inference under imperfect model conditions	Twin study on food allergy and mental health used cluster-robust SEs [19]
Explicit modeling of effect modification	Enhances interpretation of effect modification	Mixed models for pollen–pollution interaction in asthma symptom study [20]
Addressing autocorrelation and outliers in regression models	Ensures valid estimates in repeated/extreme data	GEE in asthma–pollution study; ROUT outlier removal in atherosclerosis regression [21, 22]
Justified selection of meta-analytic models	Avoids data-driven model decisions and ensures relevance	Methodological critique of heterogeneity-based model choice [23, 24]
Comprehensive reporting of heterogeneity in meta-analysis	Provides nuanced view of between-study variability	AIT meta-analysis reported I^2, Q, τ^2 and prediction interval [25, 26]
Robust implementation of Mendelian Randomization	Strengthens causal inference by addressing pleiotropy	MR-Egger, MR-PRESSO used in endocrine MR study [27]
Appropriate handling of missing data in longitudinal analysis	Improves validity and generalizability of results	Multiple imputation with FCS in pediatric asthma–allergy trajectories [28]
Verification of assumptions in time-to-event models	Ensures model assumptions are met for valid interpretation	Cox model checked via Schoenfeld and Martingale residuals [29]
Assumption awareness in exploratory methods (e.g., clustering)	Improves credibility and robustness of data-driven findings	Eczema–asthma multimorbidity study used multiple clustering strategies and robustness checks [30]

(LassoCV, logistic regression, XGBoost, and Random Forest) to rank variable importance; then selected final predictors using a Lasso regression model, showing a clear and well-documented approach [15]. However, it is important to highlight potential pitfalls when selecting independent variables without sufficient care. Problems such as multicollinearity, overfitting, or inappropriate covariate adjustment can bias results or reduce interpretability. Moreover, methods based purely on statistical criteria, such as stepwise selection, remain common but may yield unstable models, especially with small samples or correlated predictors. Statistical selection should therefore be guided by both data and domain knowledge. Moreover, complex modeling techniques such as penalized regression or machine learning approaches require sufficiently large sample sizes to ensure model stability, prevent overfitting, and yield reliable inference. In observational studies, predefined adjustment sets based on prior knowledge or causal assumptions (e.g., from DAGs) are generally preferred over automated selection based solely on statistical significance, which may lead to biased or unstable results.

Another important issue in regression models is how the relationships between variables are represented. Relying only on linear forms may overlook meaningful patterns and lead to residual confounding. For example, in a study examining oxidative balance and asthma in U.S. women, the authors used restricted cubic splines to explore possible non-linear associations and conducted subgroup analyses to investigate potential interactions. This reflects a careful approach to checking model assumptions and capturing complex relationships in the data [16].

A further relevant aspect concerns the common practice of converting continuous variables into categories, which may result in substantial information loss and reduced statistical power. In allergy-related studies, variables such as age, IgE levels, or exposure intensities are sometimes arbitrarily divided into groups, potentially masking meaningful trends, threshold effects, or nonlinear associations. This issue is not limited to regression modeling; it can also affect descriptive analyses and clinical decision-making, particularly when cut-offs are chosen without a strong empirical or biological rationale. When categorization is necessary, cut-offs should be based on clinical meaning, prior research, or statistical criteria. Common approaches include using the median, tertiles, or quartiles for balanced grouping, or applying the Youden index in diagnostic settings to optimize sensitivity and specificity. Whichever method is chosen, the rationale should be reported transparently to ensure interpretability and comparability across studies. For instance, transforming IgE concentration into “low” and “high” categories may obscure dose–response relationships or introduce artificial discontinuities. Applying flexible methods like splines allows these patterns to be modeled more accurately without discarding valuable variability in the data [17].

A helpful approach is to use tools such as DAGs to identify and adjust for confounding, as they help clarify which variables are causally related and should be accounted for in the analysis. Importantly, such tools can also help distinguish between confounders and mediators, a distinction that is often misunderstood. Confounders are variables that influence both the exposure and the outcome and must be adjusted for to obtain an unbiased estimate of the total effect. In contrast, mediators lie on the causal pathway between exposure and outcome, and adjusting for them may lead to underestimation of the total effect or introduce bias. Clarifying these roles is essential for valid causal interpretation, particularly in observational studies where randomization is absent. In a study published in the *Journal of Allergy and Clinical Immunology*, researchers constructed a DAG a priori to identify potential confounders influencing the relationship between biologic treatment and asthma exacerbations. They then applied overlap weighting to adjust for these confounders, enhancing causal inference and illustrating a structured, transparent approach to confounding control [18]. Even in randomized controlled trials, imbalances between intervention and control groups may persist due to chance, particularly in smaller samples. In such cases, covariate adjustment using regression models or weighting methods such as overlap weighting can help reduce residual confounding and improve causal interpretation. Prespecifying important prognostic covariates based on prior knowledge remains essential to avoid biased estimates. In line with this guidance, it is more informative to rely on subject-matter knowledge and prespecified causal assumptions when selecting adjustment variables. In addition to modern tools such as DAGs, classical frameworks like Hill's criteria for causality can also provide complementary guidance, particularly when interpreting associations in observational studies. While not strict rules, these criteria, such as temporality, strength of association, consistency, and biological plausibility, can help assess whether a relationship is likely to reflect a causal effect, especially in public health and epidemiologic contexts. Statistical indicators like VIFs can flag collinearity, but they should not replace controlling for confounders that are important in the context of the study. An additional consideration is the importance of using robust standard error estimators in regression analyses, especially when working with large samples. This approach improves the reliability of inference by reducing sensitivity to violations of assumptions such as homoscedasticity or independent errors. For instance, in a study on food allergy and mental health, the authors used cluster-robust standard errors to account for dependency within twin pairs, enhancing the credibility of their findings without relying on strict parametric assumptions [19]. A good example is provided by a panel study on birch pollen, air pollution, and asthma symptoms in Sweden, where the authors explicitly modeled interaction effects using mixed models to explore how air pollutant impacts varied by pollen levels. This strategy allowed them to assess whether the relationship between pollutants and respiratory outcomes changed depending on allergen exposure, rather than relying on indirect subgroup comparisons. Such an approach improves interpretability and strengthens the validity of conclusions about effect modification in complex environmental settings [20]. Mixed models more broadly deserve greater attention in allergy and immunology research, as they are well suited to the analysis of longitudinal, clustered, or hierarchical data structures frequently encountered in this field.

An additional methodological concern in regression analyses relates to potential violations of core assumptions, such as the presence of autocorrelation or influential outliers, both of which can distort inference if unaddressed. In such cases, conducting

sensitivity analyses such as comparing models with and without outliers or applying robust estimators can help evaluate the stability of findings and improve interpretability. For example, in a study on pulmonary effects of particulate matter in children with asthma, the authors employed generalized estimating equations (GEE) with an exchangeable correlation structure and robust standard errors to account for autocorrelation in repeated measurements, ensuring valid effect estimates. In another study investigating atherosclerosis regression, outliers were systematically identified and removed using the Robust Regression and Outlier Removal (ROUT) method, demonstrating careful attention to data quality and the robustness of linear modeling [21, 22].

3.5 | Statistical Assumptions in Meta-Analyses and Specialized Methods

Beyond individual-level models such as regression, meta-analysis also requires careful attention to statistical assumptions. A commonly noted issue is the use of preliminary heterogeneity tests to choose between fixed- and random-effects models. A common but suboptimal practice involves selecting between fixed- and random-effects models based on preliminary tests of heterogeneity. A better approach is to select the meta-analytic model based on the scientific rationale and study design features, rather than depending on yes-or-no heterogeneity tests. Fixed-effects models provide estimates that apply specifically to the studies included in the meta-analysis, regardless of whether the results vary across them. In contrast, random-effects models aim to generalize findings beyond the included studies, but this requires assumptions such as random sampling of studies and exchangeability, which should be clearly justified [23, 24].

Quantifying heterogeneity is also important in meta-analytical research. The I^2 statistic is frequently reported, but it is not an absolute measure of variability. It reflects the proportion of total variation attributable to between-study differences and is influenced by both within-study and between-study variance. However, I^2 alone may be insufficient, as it does not reflect the direction or magnitude of effects across studies, nor does it indicate clinical relevance. Other measures, such as τ^2 or H^2, can provide complementary information by expressing heterogeneity in different statistical terms. Each metric has its own advantages and limitations, and none should be interpreted in isolation. Rather than focusing solely on numerical indicators, the broader appraisal of inconsistency should take into account the direction and magnitude of effects across studies, as well as their potential clinical relevance. When multiple effect estimates are available from a single study (e.g., for different age groups or related outcomes), selection should follow predefined criteria based on clinical relevance, methodological consistency, or statistical independence. Preferably, the outcome most representative of the primary research question should be chosen, and sensitivity analyses may be conducted to assess the impact of alternative selections. When a single study evaluates multiple exposures in relation to the same outcome, including all of them in a meta-analysis may lead to unit-of-analysis errors due to non-independence. In such cases, either a single exposure-outcome pair should be selected based on clinical or methodological relevance, or advanced methods such as multivariate meta-analysis may be considered to account for correlations among estimates. Visual tools such as forest plots and emerging approaches that incorporate contextual thresholds can support more meaningful interpretation [25]. A good example of this broader approach to quantifying heterogeneity is provided in a meta-analysis on allergen immunotherapy for asthma prevention, where the authors reported I^2, Cochran's Q, tau-squared, and the prediction interval to comprehensively assess variability across studies, and complemented this with sensitivity analyses and publication bias assessment, although the appraisal of inconsistency was limited to statistical indicators [26].

One noteworthy method that also demands strong underlying assumptions is Mendelian Randomization, which depends on the validity of genetic instruments and the absence of horizontal pleiotropy. A clear example is provided in a study published in The Journal of Clinical Endocrinology & Metabolism, where the authors applied five complementary MR methods and explicitly used MR-Egger and MR-PRESSO to assess and correct for pleiotropy and outliers. This thorough implementation reflects good practice, as it enhances causal inference by systematically addressing key sources of bias that could otherwise distort the results [27].

An important aspect of advanced statistical analysis is the treatment of missing data, which can otherwise bias estimates and weaken validity. In a study on asthma and allergy trajectories in children, the authors assessed the extent and patterns of missing data before applying multiple imputation using fully conditional specification (FCS), a method well suited for longitudinal data. This strategy helped preserve statistical power and allowed the inclusion of individuals with partially observed data, enhancing the robustness and generalizability of the findings [28]. Beyond general model assumptions, certain methods like Cox regression have their own specific assumptions that need to be clearly tested. For example, in a study validating a pediatric asthma risk score using electronic health records, the authors applied Cox regression and explicitly tested key assumptions including proportional hazards, linearity, and the absence of outliers. They used both statistical tests and graphical diagnostics such as Schoenfeld and Martingale residuals to confirm model adequacy and ensure transparent and interpretable inference [29].

A further methodological consideration relates to exploratory approaches such as cluster analysis, which are increasingly used to identify disease patterns or generate hypotheses. While powerful for uncovering hidden data relationships, these methods rely on assumptions regarding distance metrics, clustering algorithms, and cut-off thresholds, all of which can strongly influence the resulting groupings. Principal component analysis (PCA) is another useful exploratory technique, particularly in high-dimensional datasets such as omics or biomarker studies, where it can help reduce dimensionality and reveal latent structures. A study on multimorbidity patterns in individuals with eczema and asthma provides a good example: the authors employed network and hierarchical cluster analyses based on Jaccard distance and, to assess robustness, repeated the analysis using different follow-up durations and an alternative clustering method using the Ward linkage algorithm, demonstrating a transparent and thorough approach to exploratory data analysis [30].

4 | Statistical Significance Versus Clinical Relevance

4.1 | Interpreting Statistical Significance Beyond the *p*-Value

Statistical significance, commonly assessed through *p*-values, reflects the probability that an observed effect occurred by chance. However, statistically significant findings are not necessarily meaningful from a clinical perspective. In response to widespread misunderstanding and misuse of *p*-values, the American Statistical Association (ASA) issued a statement emphasizing that *p*-values do not measure the probability that a hypothesis is true or the size or importance of an effect. The ASA cautioned against relying on arbitrary thresholds (such as $p < 0.05$) for declaring results as "statistically significant," and instead encouraged more nuanced interpretation that considers study design, context, and effect sizes [31]. Especially in large samples, even trivial differences may produce very low *p*-values, while clinically important effects may not be associated with statistical significance in smaller studies.

4.2 | Effect Sizes, Clinical Relevance, and Decision Thresholds

To enhance interpretability, researchers should also report measures of effect size, which quantify the strength or magnitude of observed associations or differences regardless of sample size. Examples include Cohen's d, Hedges' g, Phi, and Cramér's V. These metrics help contextualize whether a statistically significant finding represents a negligible or substantial effect, making it easier to judge the potential clinical or practical importance of results. They are particularly valuable when comparing findings across studies or communicating their implications to clinical audiences. Including confidence intervals for effect sizes further supports transparent reporting and allows readers to assess the precision and relevance of the estimates [32]. Key considerations for integrating statistical significance, effect sizes, and clinical relevance are summarized in Table 4.

A good example of this practice can be found in a transcriptome-wide study on allergic asthma, where independent samples t-tests were accompanied by Cohen's d to quantify the magnitude of differences between patients and controls, ensuring that statistically significant results were also evaluated in terms of practical relevance [33]. Similarly, in a study investigating hypersensitivity to environmental allergens in cats with heartworm disease, the authors reported both Cramér's V and Cohen's d to provide a complete interpretation of categorical and continuous contrasts, respectively [34]. In a methodological article on membrane protein distribution, Glass's rank biserial correlation coefficient was reported alongside nonparametric testing, further demonstrating how effect size measures can supplement significance tests to offer a more nuanced and interpretable analysis [35]. These examples illustrate how reporting effect sizes improves the interpretability of findings and supports more meaningful comparisons across research contexts. Another useful concept for interpreting the practical relevance of results is the minimally important difference (MID), which represents the smallest change in an outcome that patients perceive as beneficial and that would justify a change in

TABLE 4 | Integrating statistical significance, effect sizes, and clinical relevance in reporting.

Reporting element	Purpose and principle	Example from literature
Distinguishing statistical from clinical significance	Avoids overinterpretation of *p*-values and emphasizes real-world relevance	Methodological article recommending effect size reporting over *p*-value focus [32]
Reporting effect size measures alongside *p*-values	Quantifies the magnitude of observed effects independent of sample size	Transcriptome-wide study on asthma reported Cohen's d with t-tests [33]
Using effect sizes to interpret categorical and continuous data	Improves interpretability across data types and enhances comparative insight	Cat heartworm study reported both Cramér's V and Cohen's d [34]
Complementing nonparametric tests with effect size metrics	Provides additional context for interpreting results from nonparametric tests	Glass's rank biserial correlation coefficient used in membrane protein study [35]
Incorporating minimal clinically important differences (MCID)	Enhances patient-centered interpretation and supports clinical decision-making	Asthma study referenced MCID thresholds for ACQ, AQLQ, FEV_1 and FeNO [36]

management. Reporting whether observed effects exceed the MID helps determine not only if results are statistically significant, but also if they are meaningful in a clinical context. Including MID thresholds in the analysis allows for more patient-centered interpretation and helps translate research into practice, as long as the thresholds come from validated sources or established methods. A clear example of the application of MID is provided in an article where the authors referenced established thresholds to contextualize treatment effects in asthma. They noted that a change of 0.5 points in the Asthma Control Questionnaire (ACQ) or Asthma Quality of Life Questionnaire (AQLQ), a 10%–20% increase in FEV_1, or a 20% reduction in FeNO represent clinically meaningful improvements. Citing such MID values helps interpret whether observed changes are likely to be perceived as beneficial by patients, adding important clinical relevance to the reported outcomes [36]. While the MID provides a valuable benchmark for clinical relevance, decision-making frameworks such as GRADE (Grading of Recommendations Assessment, Development and Evaluation) recommend going further by distinguishing between multiple categories of effect magnitude, such as "trivial or none," "small but relevant," "moderate," or "large." This requires defining multiple decision thresholds to better contextualize effect sizes and support nuanced clinical recommendations. Decision thresholds refer to prespecified cut-offs that help categorize observed effects into meaningful levels of impact, facilitating interpretation aligned with clinical decision-making. This view has also been articulated by leading methodologists [37, 38]. As Gene V. Glass remarked, "Statistical significance is the least interesting thing about the results. You should describe the results in terms of measures of magnitude—not just, does a treatment affect people, but how much does it affect them [39]." Similarly, Jacob Cohen stated that "The primary product of a research inquiry is one or more measures of effect size, not *p* values [40]." These insights emphasize the importance of evaluating both the presence and the practical relevance of observed effects.

5 | Conclusions

Robust statistical methodology is essential for generating valid, interpretable, and reproducible results in allergy and immunology research. This review has highlighted key areas of good

BOX 1 | Checklist: Priorities for statistical practice in allergy and immunology research.

A. General Principles

- Clearly report statistical methods, including test purpose, assumptions, and software used
- Match statistical tests appropriately to data type and distribution
- Use informative descriptive statistics (e.g., median, IQR, proportions) consistently
- Transparently handle and report missing data

B. Study Design–Specific Considerations

For observational studies:

- Use directed acyclic graphs (DAGs) to identify potential confounders
- Apply structured strategies for confounder control

For interventional studies:

- Report and interpret effect sizes and minimally important differences (MIDs)
- Conduct and report sensitivity analyses to address assumption violations

For diagnostic studies:

- Justify statistical methods used to evaluate diagnostic accuracy
- Address potential sources of bias, including spectrum and verification bias

C. Data Type–Specific Considerations

For genetic and biomarker data:

- Apply appropriate correction for multiple comparisons
- Ensure reproducibility and transparent data preprocessing

For patient-reported outcomes (PROs):

- Use validated scoring systems and address responsiveness
- Report MID where available

D. Analytical Rigor

- Perform transparent model diagnostics in regression and machine learning
- Use interpretable methods in exploratory analyses (e.g., clustering)
- Justify all statistical analyses with appropriate references or rationale

E. Education and Reporting Standards

- Promote statistical education among researchers and clinicians
- Encourage stronger statistical peer review in journals
- Provide clear guidance for selecting, justifying, and reporting statistical methods

BOX 2 | Major milestone discoveries.

Box 2 requirement (major milestone discoveries) is addressed through Tables 1–4, which compile well-documented examples of statistical good practice across key methodological areas in allergy and immunology research.

practice, including transparent reporting, careful verification of assumptions, and the interpretation of results beyond *p*-values. Published examples from recent allergy and immunology studies illustrate how aligning statistical methods with study design enhances scientific credibility. Greater emphasis should be placed on clear documentation of analytical choices and the rationale behind them, particularly in the context of model selection and data characteristics. Researchers are encouraged to use effect size measures and clinically meaningful thresholds to support more nuanced interpretation. Good statistical practice not only strengthens the evidence base but also facilitates communication with clinical audiences. There remains a need for broader implementation of methodological standards across the publication process. Advancing statistical transparency and rigor will ultimately contribute to higher quality research and better-informed decisions in clinical and public health settings. To support implementation, future research perspectives and milestone discoveries are summarized for each thematic section in tabular format, ensuring contextual clarity and practical relevance. While rigorous statistical methodology is essential for generating valid and reproducible results, it is equally important not to lose sight of the underlying biological and immunological questions. Preliminary or exploratory findings, even if based on limited data, can provide valuable insights and guide future research directions. Therefore, statistical sophistication should serve to complement and clarify, rather than obscure, the scientific and clinical significance of study outcomes Box 1 and 2.

Author Contributions

All authors contributed to the conception, structure, development, and review of the manuscript.

Acknowledgements

The authors wishes to thank Prof. Cezmi A. Akdis, Editor-in-Chief of Allergy, for the kind invitation to contribute this review article.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

1. G. D'Arrigo, S. A. El Hafeez, S. Mezzatesta, et al., "Common Mistakes in Biostatistics," *Clinical Kidney Journal* 17 (2024): sfae197.

2. D. H. Kim, S. G. Kim, and Kwak, "Comprehensive Guidelines for Appropriate Statistical Analysis Methods in Research," *Korean Journal of Anesthesiology* 77 (2024): 503–517.

3. M. Ordak, "Poor Statistical Reporting: Do We Have a Reason for Concern? A Narrative Review and Recommendations," *Current Opinion in Allergy and Clinical Immunology* 24 (2024): 237–242.

4. M. Ordak, "Statistical Reviews in Journals of the World Association of Medical Editors," *Polish Archives of Internal Medicine* 134 (2024): 16778.

5. E. Pacheco Da Silva, T. Weinmann, J. Gerlich, et al., "Exposure Profiles for the Long-Term Use of Disinfectants and Cleaning Products and Asthma," *Allergy* 80 (2025): 996–1005.

6. A. Lemoine, N. Kapel, I. Nicolis, et al., "Identification of Gut Biomarkers of FPIES in a Longitudinal Comparative Pediatric Study," *Allergy* 80 (2025): 1389–1399.

7. S. Hirai, K. Yamamoto-Hanada, K. Pak, M. Saito-Abe, T. Fukuie, and Y. Ohya, "Predictive Modeling for Cow's Milk Allergy Remission by Low-Dose Oral Immunotherapy in Young Children," *World Allergy Organization Journal* 17 (2024): 100910.

8. C. Dong, S. Lu, Z. Deng, et al., "Role for Complement C5 in Eosinophilic Inflammation of Severe Asthma," *Allergy* (2025), https://doi.org/10.1111/all.16616. Epub ahead of print.

9. J. Oh, M. Lee, M. Kim, et al., "Incident Allergic Diseases in Post-COVID-19 Condition: Multinational Cohort Studies From South Korea, Japan and the UK," *Nature Communications* 15 (2024): 2830.

10. P. R. C. Macedo, P. Moraes, L. K. Arruda, F. F. M. Castro, J. Kalil, and C. E. S. Galvão, "Efficacy and Safety of Sublingual Immunotherapy Using a Combination of Dermatophagoides Pteronyssinus and Blomia Tropicalis Extracts in Patients With Allergic Rhinitis: A Randomized, Double-Blind, Placebo-Controlled Trial," *World Allergy Organization Journal* 18 (2025): 101020.

11. T. Lumley, P. Diehr, S. Emerson, and L. Chen, "The Importance of the Normality Assumption in Large Public Health Data Sets," *Annual Review of Public Health* 23 (2002): 151–169.

12. I. Randhawa and N. Marsteller, "Long-Term Efficacy and Safety of Cow's Milk Anaphylaxis Specific Immunotherapy: Allergen Unresponsiveness via the Tolerance Induction Program," *Journal of Allergy and Clinical Immunology: Global* 3 (2024): 100285.

13. Z. D. Draelos, "A Clinical Evaluation of a Permanent Hair Dye Designed to Reduce Allergic Contact Dermatitis and Hair Damage," *Journal of Cosmetic Dermatology* 21 (2022): 3925–3928.

14. K. Jonsson, M. Barman, S. Moberg, et al., "Fat Intake and Breast Milk Fatty Acid Composition in Farming and Nonfarming Women and Allergy Development in the Offspring," *Pediatric Research* 79 (2016): 114–123.

15. J. Li, M. Y. Zhou, Y. Li, et al., "Clinical Prediction Model by Machine Learning to Determine the Results of Maternal Dietary Avoidance in Food Protein-Induced Allergic Proctocolitis Infants," *Frontiers in Pediatrics* 13 (2025): 1612076.

16. X. Qi, T. Zhou, and J. Tang, "Correlations Between Oxidative Balance Score and Female Asthma Among U.S. Adults," *Scientific Reports* 14 (2024): 22451.

17. J. Gauthier, Q. V. Wu, and T. A. Gooley, "Cubic Splines to Model Relationships Between Continuous Variables and Outcomes: A Guide for Clinicians," *Bone Marrow Transplantation* 55 (2020): 675–680.

18. A. T. Akenroye, J. B. Segal, G. Zhou, et al., "Comparative Effectiveness of Omalizumab, Mepolizumab, and Dupilumab in Asthma: A Target Trial Emulation," *Journal of Allergy and Clinical Immunology* 151 (2023): 1269–1276.

19. H. Karim, C. Lundholm, T. Gong, B. Brew, M. Silverman, and C. Almqvist, "Food Allergy and Mental Health in Children and Adolescents – The Role of Shared Familial Environment," *Clinical and Experimental Allergy* 55 (2025): 175–186.

20. H. K. Carlsen, S. L. Haga, D. Olsson, et al., "Birch Pollen, Air Pollution and Their Interactive Effects on Airway Symptoms and Peak

Expiratory Flow in Allergic Asthma During Pollen Season – A Panel Study in Northern and Southern Sweden," *Environmental Health* 21 (2022): 63.

21. J. Q. Koenig, T. F. Mar, R. W. Allen, et al., "Pulmonary Effects of Indoor- and Outdoor-Generated Particles in Children With Asthma," *Environmental Health Perspectives* 113 (2005): 499–503.

22. D. M. Boucher, S. Robichaud, V. Lorant, et al., "Age-Related Impairments in Immune Cell Efferocytosis and Autophagy Hinder Atherosclerosis Regression," *Arteriosclerosis, Thrombosis, and Vascular Biology* 45 (2025): 481–495.

23. K. Rice, J. P. T. Higgins, and T. Lumley, "A re-Evaluation of Fixed Effect(s) Meta-Analysis," *Journal of the Royal Statistical Society, Series A* 181 (2018): 205.

24. J. P. T. Higgins, S. G. Thompson, and D. J. Spiegelhalter, "A re-Evaluation of Random-Effects Meta-Analysis," *Journal of the Royal Statistical Society, Series A* 172 (2009): 137–159.

25. M. Ordak, G. W. Canonica, G. Paoletti, L. Brussino, D. Carvalho, and D. Di Bona, "Meta-Analysis in Allergy – Statistical Recommendations," *Allergy* 79 (2024): 1616–1618.

26. M. Farraia, I. Paciência, F. Castro Mendes, et al., "Allergen Immunotherapy for Asthma Prevention: A Systematic Review and Meta-Analysis of Randomized and Non-Randomized Controlled Studies," *Allergy* 77 (2022): 1719–1735.

27. Y. Chen, X. Chen, and Z. Zhang, "Association Between Immune Cells and Urticaria: A Bidirectional Mendelian Randomization Study," *Frontiers in Immunology* 15 (2024): 1439331.

28. D. Lisik, G. Wennergren, H. Kankaanranta, et al., "Asthma and Allergy Trajectories in Children Based on Combined Parental Report and Register Data," *Pediatric Allergy and Immunology* 35 (2024): e14254.

29. A. H. Owora, B. Jiang, Y. Shah, B. Gaston, and M. Boustani, "External Validation and Update of the Pediatric Asthma Risk Score as a Passive Digital Marker for Childhood Asthma Using Integrated Electronic Health Records," *EClinicalMedicine* 84 (2025): 103254.

30. A. R. Mulick, A. D. Henderson, D. Prieto-Merino, et al., "Novel Multimorbidity Clusters in People With Eczema and Asthma: A Population-Based Cluster Analysis," *Scientific Reports* 12 (2022): 21866.

31. R. L. Wasserstein and N. A. Lazar, "The ASA Statement on *p*-Values: Context, Process, and Purpose," *American Statistician* 70 (2016): 129–133.

32. M. Ordak, "Multiple Comparisons and Effect Size: Statistical Recommendations for Authors Planning to Submit an Article to Allergy," *Allergy* 78 (2023): 1145–1147.

33. F. Cui, R. Liu, L. Wang, J. He, and Y. Xu, "Identification and Validation of Differentially Expressed Genes in Allergic Asthma Pathogenesis Using Whole-Transcriptome Sequencing," *Frontiers in Medicine* 12 (2025): 1545095.

34. S. N. García-Rodríguez, N. Costa-Rodríguez, J. I. Matos, et al., "Feline Heartworm Disease and Environmental Allergens Hypersensitivity: Is There a Link?," *Parasites & Vectors* 16 (2023): 192.

35. D. Guerrero-Given, S. L. Goldin, C. I. Thomas, S. A. Anthony, D. Jerez, and N. Kamasawa, "Gold in-and-Out: A Toolkit for Analyzing Subcellular Distribution of Immunogold-Labeled Membrane Proteins in Freeze-Fracture Replica Images," *Frontiers in Neuroanatomy* 16 (2022): 855218.

36. F. E. van Boven, G. J. Braunstahl, L. R. Arends, et al., "House Dust Mite Allergen Avoidance Strategies for the Treatment of Allergic Asthma: A Hypothesis-Generating Meta-Analysis," *World Allergy Organization Journal* 17 (2024): 100919.

37. H. J. Schünemann, I. Neumann, M. Hultcrantz, et al., "GRADE Guidance 35: Update on Rating Imprecision for Assessing Contextualized Certainty of Evidence and Making Decisions," *Journal of Clinical Epidemiology* 150 (2022): 225–242.

38. G. P. Morgano, W. Wiercioch, D. Piovani, et al., "Defining Decision Thresholds for Judgments on Health Benefits and Harms Using the GRADE Evidence to Decision (EtD) Frameworks: A Randomized Methodological Study (GRADE-THRESHOLD)," *Journal of Clinical Epidemiology* 179 (2025): 111639.

39. R. B. Kline, *Beyond Significance Testing Reforming Data Analysis Methods in Behavioral Research* (American Psychological Association, 2004).

40. J. Cohen, "Things I Have Learned (So Far)," *American Psychologist* 45 (1990): 1304–1312.