

Table S1. Articles used for identifying statistical issues in the statistical analysis section

PMID	Title
25848638	Important statistical points to improve and promote the methodology of the articles on medical sciences, particularly nephrology and kidney; a review article.
19260441	Statistical pitfalls in medical research.
17299669	Statistical errors in medical research - a review of common pitfalls.
22125459	Misuse of statistical methods in 10 leading Chinese medical journals in 1998 and 2008.
19435635	Common statistical and research design problems in manuscripts submitted to high-impact psychiatry journals: what editors and reviewers want authors to know.
11929408	Statistical issues--significantly important in medical research.
29209453	Statistical methods and common problems in medical or biomedical science research
26256144	Statistical Considerations for Readers and Authors of the Forensic Science Literature.
33344343	Pitfalls in Article Submissions for Publication.
21854631	Common statistical and research design problems in manuscripts submitted to high-impact medical journals
33948238	Learning gaps among statistical competencies for clinical and translational science learners.
30881471	The Principles of Biomedical Scientific Writing: Materials and Methods.

PMID: PubMed ID; Search terms used: Statistical “problems or issues or pitfalls or errors or misuse” in medical research.

Table S2. Biostatistics applications at different phases of research design

Research steps	Goal	Biostatistics Applications
Research question	Refining questions by converting good questions to great questions	Exposure, primary and secondary outcomes and their forms in view of hypothesis, multiplicity, multiple comparisons, and study objectives
Study design	Optimum design	Eliminating biases by setting design features at the beginning and during the study execution
Data generation process	Selection of subjects and measurement process to address hypothesis/question	Assist in probabilistic/non-probabilistic procedures for selecting subjects, the optimum number of measurements in different conditions, outcome distribution consideration, and other factors.
Sample size & power analyses	Optimum study size	Sample size simulations and computation
Factors and collection	Avoiding bad data	List of measurements on the conceptual framework, their forms and roles for statistical models
Data management and analytic dataset	Obtain reusable datasets with data transparency	Minimizing data errors by data checks, coding, proper labeling and preparing the analytic dataset

Data analysis	Obtain meaningful findings	Conduct and report necessary steps for measuring accuracy, efficiency, reliability, consistency, and applicability
Interpreting data	Obtain meaningful conclusions	Avoiding incorrect or over-interpretation not aligned with effect size, confidence interval, p-value, study design, protocol deviation, and consistencies and inconsistencies between findings from primary, secondary, and sensitivity analyses
Presenting data	Transparent message and interpretation	Assist in providing transparent reporting, usability, and meaningful interpretation of data

Table S3. The evidence-based preferred methods in different conditions/settings

Condition	EBB preferred method	Standard or common practice
Unpaired data	Welch t-test or bootstrap t-test	Unpaired t-test
Discrete or ordinal data	Bootstrap t-test or nonparametric test	Parametric test
Pre-post changes between groups	ANCOVA	Unpaired or paired t-tests or percent change
Propensity matched	Paired analysis	Unpaired analysis
Binary data in a cohort or clinical trial	Relative risk models	Odds ratio models
Binary data in a case-control study	Odds ratio models	Odds ratio models
Binary data in a cross-sectional study	Relative risk models	Odds ratio models
Clustered interventional data	Robust GEE	GLM
Longitudinal data	GEE/GLMM	GLM
Time to event data	Cox/PM survival models	Logistic regression
Diagnostic study data	Robust multinomial logistic /GEE	Unadjusted analysis
Non-linearity assessment	Multivariable polynomial	Transformation
Interventional study	Alternative of p-values	Nominal p-value
Nonrandomized interventional study	Propensity score-adjusted	Simple adjusted analysis
Exploratory objective study	Backward variable selection	Stepwise
Association objective study	Full adjustment of all relevant variables	Univariate selection with $p < 0.05$
Prediction objective study	Random forest/bootstrap/backward	Univariate selection with $p < 0.05$
Highly skewed outcome data	Generalized nonlinear model	Linear regression on transformed data
Count data	Count models/inflated count model	Logistic regression after ignoring count structure

EBB: evidence-based biostatistics; ANCOVA: analysis of covariance; GEE; generalized estimating equations; GLMM: generalized linear mixed model; PM: parametric; GLM: generalized linear model.