How to write statistical analysis section in medical research

Alok Kumar Dwivedi [ORCID]

ABSTRACT

Reporting of statistical analysis is essential in any clinical and translational research study. However, medical research studies sometimes report statistical analysis that is either inappropriate or insufficient to attest to the accuracy and validity of findings and conclusions. Published works involving inaccurate statistical analyses and insufficient reporting influence the conduct of future scientific studies, including meta-analyses and medical decisions. Although the biostatistical practice has been improved over the years due to the involvement of statistical reviewers and collaborators in research studies, there remain areas of improvement for transparent reporting of the statistical analysis section in a study. Evidence-based biostatistics practice throughout the research is useful for generating reliable data and translating meaningful data to meaningful interpretation and decisions in medical research. Most existing research reporting guidelines do not provide guidance for reporting methods in the statistical analysis section that helps in evaluating the quality of findings and data interpretation. In this report, we highlight the global and critical steps to be reported in the statistical analysis of grants and research articles. We provide clarity and the importance of understanding study objective types, data generation process, effect size use, evidence-based biostatistical methods use, and development of statistical models through several thematic frameworks. We also provide published examples of adherence or non-adherence to methodological standards related to each step in the statistical analysis and their implications. We believe the suggestions provided in this report can have far-reaching implications for education and strengthening the quality of statistical reporting and biostatistical practice in medical research.

INTRODUCTION

Biostatistics is the overall approach to how we realistically and feasibly execute a research idea to produce meaningful data and translate data to meaningful interpretation and decisions. In this era of evidence-based medicine and practice, basic biostatistical knowledge becomes essential for critically appraising research articles and implementing findings for better patient management, improving healthcare, and research planning.[1] However, it may not be sufficient for the proper execution and reporting of statistical analyses in studies.[2][3] Three things are required for statistical analyses, namely knowledge of the conceptual framework of variables, research design, and evidence-based applications of statistical analysis with statistical software.[4][5] The conceptual framework provides possible biological and clinical pathways between independent variables and outcomes with role specification of variables. The research design provides a protocol of study design and data generation process (DGP), whereas the evidence-based statistical analysis approach provides guidance for selecting and implementing approaches after evaluating data with the research design.[2][5] Ocaña-Riola[6] reported a substantial percentage of articles from high-impact medical journals contained errors in statistical analysis or data interpretation. These errors in statistical analyses and interpretation of results do not only impact the reliability of research findings but also influence the medical decision-making and planning and execution of other related studies. A survey of consulting biostatisticians in the USA reported that researchers frequently request biostatisticians for performing inappropriate statistical analyses and inappropriate reporting of data.[7] This implies that there is a need to enforce standardized reporting of the statistical analysis section in medical research which can also help rreviewers and investigators to improve the methodological standards of the study.

Biostatistical practice in medicine has been improving over the years due to continuous efforts in promoting awareness and involving expert services on biostatistics, epidemiology, and research design in clinical and translational research.[8–11] Despite these efforts, the quality of reporting of statistical analysis in research studies has often been suboptimal.[12][13] We noticed that none of the methods reporting documents were developed using evidence-based biostatistics (EBB) theory and practice. The EBB practice implies that the selection of statistical analysis methods for statistical analyses and the steps of results reporting and interpretation should be grounded based on the evidence generated in the scientific literature and according to the study objective type and design.[5] Previous works have not properly elucidated the importance of understanding EBB concepts and related reporting in the write-up

BMJ

of statistical analyses. As a result, reviewers sometimes ask to present data or execute analyses that do not match the study objective type.[14] We summarize the statistical analysis steps to be reported in the statistical analysis section based on review and thematic frameworks.

METHODS

We identified articles describing statistical reporting problems in medicine using different search terms (online supplemental table 1). Based on these studies, we prioritized commonly reported statistical errors in analytical strategies and developed essential components to be reported in the statistical analysis section of research grants and studies. We also clarified the purpose and the overall implication of reporting each step in statistical analyses through various examples.

RESULTS

Although biostatistical inputs are critical for the entire research study (online supplemental table 2), biostatistical consultations were mostly used for statistical analyses only[15]. Even though the conduct of statistical analysis mismatched with the study objective and DGP was identified as the major problem in articles submitted to high-impact medical journals.[16] In addition, multivariable analyses were often inappropriately conducted and reported in published studies.[17 18] In light of these statistical errors, we describe

the reporting of the following components in the statistical analysis section of the study.

Step 1: specify study objective type and outcomes (overall approach)

The study objective type provides the role of important variables for a specified outcome in statistical analyses and the overall approach of the model building and model reporting steps in a study. In the statistical framework, the problems are classified into descriptive and inferential/analytical/confirmatory objectives. In the epidemiological framework, the analytical and prognostic problems are broadly classified into association, explanatory, and predictive objectives.[19] These study objectives (figure 1) may be classified into six categories: (1) exploratory, (2) association, (3) causal, (4) intervention, (5) prediction and (6) clinical decision models in medical research.[20]

The exploratory objective type is a specific type of determinant study and is commonly known as risk factors or correlates study in medical research. In an exploratory study, all covariates are considered equally important for the outcome of interest in the study. The goal of the exploratory study is to present the results of a model which gives higher accuracy after satisfying all model-related assumptions. In the association study, the investigator identifies predefined exposures of interest for the outcome, and variables other than exposures are also important for the interpretation

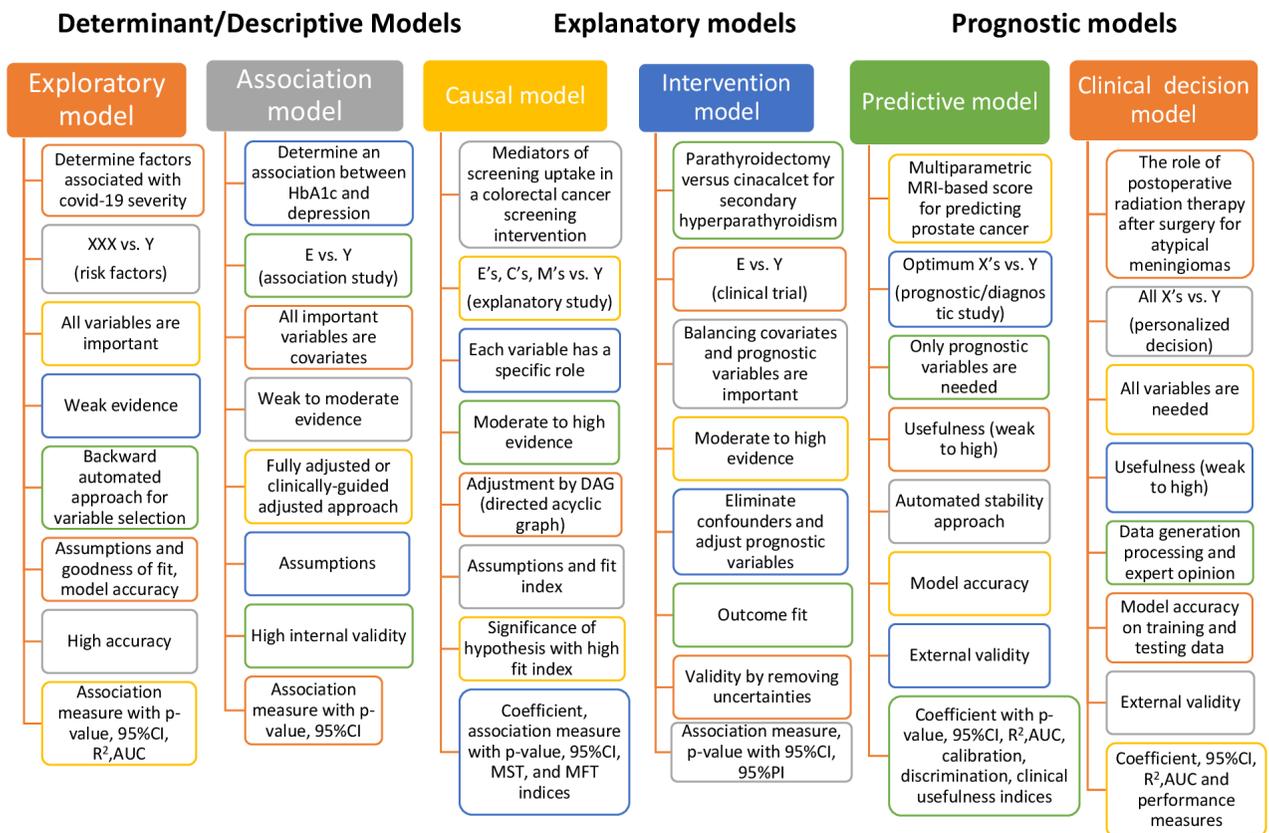**Figure 1** Comparative assessments of developing and reporting of study objective types and models. Association measures include odds ratio, risk ratio, or hazard ratio. AUC, area under the curve; C, confounder; CI, confidence interval; E, exposure; HbA1C: hemoglobin A1c; M, mediator; MFT, model fit test; MST, model specification test; PI, predictive interval; R^2, coefficient of determinant; X, independent variable; Y, outcome.

and considered as covariates. The goal of an association study is to present the adjusted association of exposure with outcome.[20] In the causal objective study, the investigator is interested in determining the impact of exposure(s) on outcome using the conceptual framework. In this study objective, all variables should have a predefined role (exposures, confounders, mediators, covariates, and predictors) in a conceptual framework. A study with a causal objective is known as an explanatory or a confirmatory study in medical research. The goal is to present the direct or indirect effects of exposure(s) on an outcome after assessing the model's fitness in the conceptual framework.[19 21] The objective of an interventional study is to determine the effect of an intervention on outcomes and is often known as randomized or non-randomized clinical trials in medical research. In the intervention objective model, all variables other than the intervention are treated as nuisance variables for primary analyses. The goal is to present the direct effect of the intervention on the outcomes by eliminating biases.[22–24] In the predictive study, the goal is to determine an optimum set of variables that can predict the outcome, particularly in external settings. The clinical decision models are a special case of prognostic models in which high dimensional data at various levels are used for risk stratification, classification, and prediction. In this model, all variables are considered input features. The goal is to present a decision tool that has high accuracy in training, testing, and validation data sets.[20 25] Biostatisticians or applied researchers should properly discuss the intention of the study objective type before proceeding with statistical analyses. In addition, it would be a good idea to prepare a conceptual model framework regardless of study objective type to understand study concepts.

Examples

A study[26] showed a favorable effect of the beta-blocker intervention on survival outcome in patients with advanced human epidermal growth factor receptor (HER2)-negative breast cancer without adjusting for all the potential confounding effects (age or menopausal status and Eastern Cooperative Oncology Performance Status) in primary analyses or validation analyses or using a propensity score-adjusted analysis, which is an EBB preferred method for analyzing non-randomized studies.[27] Similarly, another study had the goal of developing a predictive model for prediction of Alzheimer's disease progression.[28] However, this study did not internally or externally validate the performance of the model as per the requirement of a predictive objective study. In another study,[29] investigators were interested in determining an association between metabolic syndrome and hepatitis C virus. However, the authors did not clearly specify the outcome in the analysis and produced conflicting associations with different analyses.[30] Thus, the outcome should be clearly specified as per the study objective type.

Step 2: specify effect size measure according to study design (interpretation and practical value)

The study design provides information on the selection of study participants and the process of data collection conditioned on either exposure or outcome (figure 2). The appropriate use of effect size measure, tabular presentation of results, and the level of evidence are mostly determined by the study design.[31 32] In cohort or clinical trial study designs, the participants are selected based on exposure status and are followed up for the development
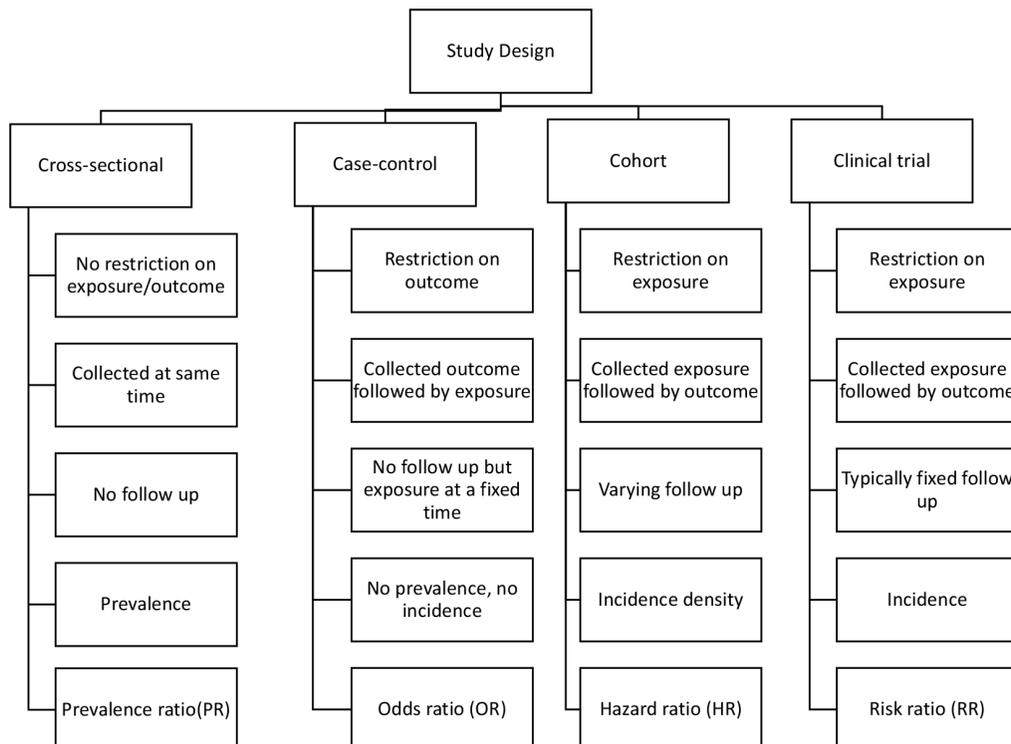

Figure 2 Effect size according to study design.

of the outcome. These study designs can provide multiple outcomes, produce incidence or incidence density, and are preferred to be analyzed with risk ratio (RR) or hazards models. In a case–control study, the selection of participants is conditioned on outcome status. This type of study can have only one outcome and is preferred to be analyzed with an odds ratio (OR) model. In a cross-sectional study design, there is no selection restriction on outcomes or exposures. All data are collected simultaneously and can be analyzed with a prevalence ratio model, which is mathematically equivalent to the RR model.[33] The reporting of effect size measure also depends on the study objective type. For example, predictive models typically require reporting of regression coefficients or weight of variables in the model instead of association measures, which are required in other objective types. There are agreements and disagreements between OR and RR measures. Due to the constancy and symmetricity properties of OR, some researchers prefer to use OR in studies with common events. Similarly, the collapsibility and interpretability properties of RR make it more appealing to use in studies with common events.[34] To avoid variable practice and interpretation issues with OR, it is recommended to use RR models in all studies except for case–control and nested case–control studies, where OR approximates RR and thus OR models should be used. Otherwise, investigators may report sufficient data to compute any ratio measure. Biostatisticians should educate investigators on the proper interpretation of ratio measures in the light of study design and their reporting.[34 35]

Examples
Investigators sometimes either inappropriately label their study design[36 37] or report effect size measures not aligned with the study design,[38 39] leading to difficulty in results interpretation and evaluation of the level of evidence. The proper labeling of study design and the appropriate use of effect size measure have substantial implications for results interpretation, including the conduct of systematic review and meta-analysis.[40] A study[31] reviewed the frequency of reporting OR instead of RR in cohort studies and randomized clinical trials (RCTs) and found that one-third of the cohort studies used an OR model, whereas 5% of RCTs used an OR model. The majority of estimated ORs from these studies had a 20% or higher deviation from the corresponding RR.

Step 3: specify study hypothesis, reporting of p values, and interval estimates (interpretation and decision)
The clinical hypothesis provides information for evaluating formal claims specified in the study objectives, while the statistical hypothesis provides information about the population parameters/statistics being used to test the formal claims. The inference about the study hypothesis is typically measured by p value and confidence interval (CI). A smaller p value indicates that the data support against the null hypothesis. Since the p value is a conditional probability, it can never tell about the acceptance or rejection of the null hypothesis. Therefore, multiple alternative strategies of p values have been proposed to strengthen the credibility of conclusions.[41 42] Adaption of these alternative strategies is only needed in the explanatory objective studies. Although

exact p values are recommended to be reported in research studies, p values do not provide any information about the effect size. Compared with p values, the CI provides a confidence range of the effect size that contains the true effect size if the study were repeated and can be used to determine whether the results are statistically significant or not.[43] Both p value and 95% CI provide complementary information and thus need to be specified in the statistical analysis section.[24 44]

Researchers often test one or more comparisons or hypotheses. Accordingly, the side and the level of significance for considering results to be statistically significant may change. Furthermore, studies may include more than one primary outcome that requires an adjustment in the level of significance for multiplicity. All studies should provide the interval estimate of the effect size/regression coefficient in the primary analyses. Since the interpretation of data analysis depends on the study hypothesis, researchers are required to specify the level of significance along with the side (one-sided or two-sided) of the p value in the test for considering statistically significant results, adjustment of the level of significance due to multiple comparisons or multiplicity, and reporting of interval estimates of the effect size in the statistical analysis section.[45]

Examples
A study[46] showed a significant effect of fluoxetine on relapse rates in obsessive-compulsive disorder based on a one-sided p value of 0.04. Clearly, there was no reason for using a one-sided p value as opposed to a two-sided p value. A review of the appropriate use of multiple test correction methods in multiarm clinical trials published in major medical journals in 2012 identified over 50% of the articles did not perform multiple-testing correction.[47] Similar to controlling a familywise error rate due to multiple comparisons, adjustment of the false discovery rate is also critical in studies involving multiple related outcomes. A review of RCTs for depression between 2007 and 2008 from six journals reported that only limited studies (5.8%) accounted for multiplicity in the analyses due to multiple outcomes.[48]

Step 4: account for DGP in the statistical analysis (accuracy)
The study design also requires the specification of the selection of participants and outcome measurement processes in different design settings. We referred to this specific design feature as DGP. Understanding DGP helps in determining appropriate modeling of outcome distribution in statistical analyses and setting up model premises and units of analysis.[4] DGP (figure 3) involves information on data generation and data measures, including the number of measurements after random selection, complex selection, consecutive selection, pragmatic selection, or systematic selection. Specifically, DGP depends on a sampling setting (participants are selected using survey sampling methods and one subject may represent multiple participants in the population), clustered setting (participants are clustered through a recruitment setting or hierarchical setting or multiple hospitals), pragmatic setting (participants are selected through mixed approaches), or systematic review setting (participants are selected from published studies).
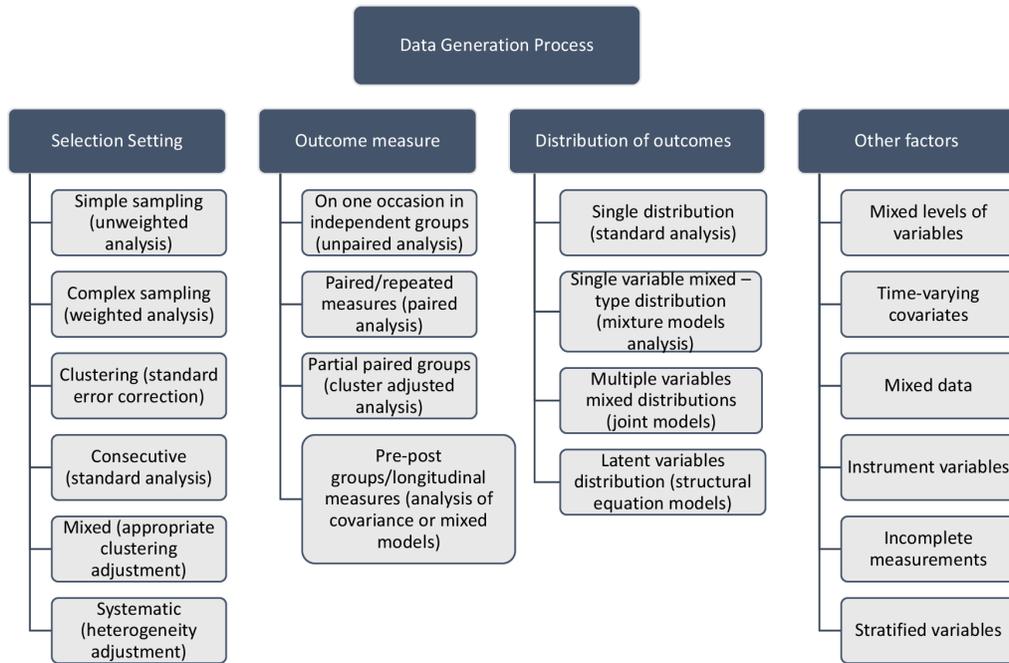
Figure 3 Common features of the data generation process.

DGP also depends on the measurements of outcomes in an unpaired setting (measured on one occasion only in independent groups), paired setting (measured on more than one occasion or participants are matched on certain subject characteristics), or mixed setting (measured on more than one occasion but interested in comparing independent groups). It also involves information regarding outcomes or exposure generation processes using quantitative or categorical variables, quantitative values using labs or validated instruments, and self-reported or administered tests yielding a variety of data distributions, including individual distribution, mixed-type distribution, mixed distributions, and latent distributions. Due to different DGPs, study data may include messy or missing data, incomplete/partial measurements, time-varying measurements, surrogate measures, latent measures, imbalances, unknown confounders, instrument variables, correlated responses, various levels of clustering, qualitative data, or mixed data outcomes, competing events, individual and higher-level variables, etc. The performance of statistical analysis, appropriate estimation of standard errors of estimates and subsequently computation of p values, the generalizability of findings, and the graphical display of data rely on DGP. Accounting for DGP in the analyses requires proper communication between investigators and biostatisticians about each aspect of participant selection and data collection, including measurements, occasions of measurements, and instruments used in the research study.

Examples

A study[49] compared the intake of fresh fruit and komatsuna juice with the intake of commercial vegetable juice on metabolic parameters in middle-aged men using an RCT. The study was criticized for many reasons, but primarily for incorrect statistical methods not aligned with the study DGP.[50] Similarly, another study[51] highlighted that 80% of

published studies using the Korean National Health and Nutrition Examination Survey did not incorporate survey sampling structure in statistical analyses, producing biased estimates and inappropriate findings. Likewise, another study[52] highlighted the need for maintaining methodological standards while analyzing data from the National Inpatient Sample. A systematic review[53] identified that over 50% of studies did not specify whether a paired t-test or an unpaired t-test was performed in statistical analysis in the top 25% of physiology journals, indicating poor transparency in reporting of statistical analysis as per the data type. Another study[54] also highlighted the data displaying errors not aligned with DGP. As per DGP, delay in treatment initiation of patients with cancer defined from the onset of symptom to treatment initiation should be analyzed into three components: patient/primary delay, secondary delay, and tertiary delay.[55] Similarly, the number of cancerous nodes should be analyzed with count data models.[56] However, several studies did not analyze such data according to DGP.[57 58]

Step 5: apply EBB methods specific to study design features and DGP (efficiency and robustness)

The continuous growth in the development of robust statistical methods for dealing with a specific problem produced various methods to analyze specific data types. Since multiple methods are available for handling a specific problem yet with varying performances, heterogeneous practices among applied researchers have been noticed. Variable practices could also be due to a lack of consensus on statistical methods in literature, unawareness, and the unavailability of standardized statistical guidelines.[2 5 59] However, it becomes sometimes difficult to differentiate whether a specific method was used due to its robustness, lack of awareness, lack of accessibility of statistical software

to apply an alternative appropriate method, intention to produce expected results, or ignorance of model diagnostics. To avoid heterogeneous practices, the selection of statistical methodology and their reporting at each stage of data analysis should be conducted using methods according to EBB practice.[5] Since it is hard for applied researchers to optimally select statistical methodology at each step, we encourage investigators to involve biostatisticians at the very early stage in basic, clinical, population, translational, and database research. We also appeal to biostatisticians to develop guidelines, checklists, and educational tools to promote the concept of EBB. As an effort, we developed the statistical analysis and methods in biomedical research (SAMBR) guidelines for applied researchers to use EBB methods for data analysis.[5] The EBB practice is essential for applying recent cutting-edge robust methodologies to yield accurate and unbiased results. The efficiency of statistical methodologies depends on the assumptions and DGP. Therefore, investigators may attempt to specify the choice of specific models in the primary analysis as per the EBB.

Examples

Although details of evidence-based preferred methods are provided in the SAMBR checklists for each study design/objective,[5] we have presented a simplified version of evidence-based preferred methods for common statistical analysis (online supplemental table 3). Several examples are available in the literature where inefficient methods not according to EBB practice have been used.[31 57 60]

Step 6: report variable selection method in the multivariable analysis according to study objective type (unbiased)

Multivariable analysis can be used for association, prediction or classification or risk stratification, adjustment, propensity score development, and effect size estimation.[61] Some biological, clinical, behavioral, and environmental factors may directly associate or influence the relationship between exposure and outcome. Therefore, almost all health studies require multivariable analyses for accurate and unbiased interpretations of findings (figure 1). Analysts should develop an adjusted model if the sample size permits. It is a misconception that the analysis of RCT does not require adjusted analysis. Analysis of RCT may require adjustment for prognostic variables.[23] The foremost step in model building is the entry of variables after finalizing the appropriate parametric or non-parametric regression model. In the exploratory model building process due to no preference of exposures, a backward automated approach after including any variables that are significant at 25% in the unadjusted analysis can be used for variable selection.[62 63] In the association model, a manual selection of covariates based on the relevance of the variables should be included in a fully adjusted model.[63] In a causal model, clinically guided methods should be used for variable selection and their adjustments.[20] In a non-randomized interventional model, efforts should be made to eliminate confounding effects through propensity score methods and the final propensity score-adjusted multivariable model may adjust any prognostic variables, while a randomized study simply should adjust any prognostic variables.[27] Maintaining the

event per variable (EVR) is important to avoid overfitting in any type of modeling; therefore, screening of variables may be required in some association and explanatory studies, which may be accomplished using a backward stepwise method that needs to be clarified in the statistical analyses.[10] In a predictive study, a model with an optimum set of variables producing the highest accuracy should be used. The optimum set of variables may be screened with the random forest method or bootstrap or machine learning methods.[64 65] Different methods of variable selection and adjustments may lead to different results. The screening process of variables and their adjustments in the final multivariable model should be clearly mentioned in the statistical analysis section.

Examples

A study[66] evaluating the effect of hydroxychloroquine (HDQ) showed unfavorable events (intubation or death) in patients who received HDQ compared with those who did not (hazard ratio (HR): 2.37, 95% CI 1.84 to 3.02) in an unadjusted analysis. However, the propensity score-adjusted analyses as appropriate with the interventional objective model showed no significant association between HDQ use and unfavorable events (HR: 1.04, 95% CI 0.82 to 1.32), which was also confirmed in multivariable and other propensity score-adjusted analyses. This study clearly suggests that results interpretation should be based on a multivariable analysis only in observational studies if feasible. A recent study[10] noted that approximately 6% of multivariable analyses based on either logistic or Cox regression used an inappropriate selection method of variables in medical research. This practice was more commonly noted in studies that did not involve an expert biostatistician. Another review[61] of 316 articles from high-impact Chinese medical journals revealed that 30.7% of articles did not report the selection of variables in multivariable models. Indeed, this inappropriate practice could have been identified more commonly if classified according to the study objective type.[18] In RCTs, it is uncommon to report an adjusted analysis based on prognostic variables, even though an adjusted analysis may produce an efficient estimate compared with an unadjusted analysis. A study assessing the effect of preemptive intervention on development outcomes showed a significant effect of an intervention on reducing autism spectrum disorder symptoms.[67] However, this study was criticized by Ware[68] for not reporting non-significant results in unadjusted analyses. If possible, unadjusted estimates should also be reported in any study, particularly in RCTs.[23 68]

Step 7: provide evidence for exploring effect modifiers (applicability)

Any variable that modifies the effect of exposure on the outcome is called an effect modifier or modifier or an interacting variable. Exploring the effect modifiers in multivariable analyses helps in (1) determining the applicability/generalizability of findings in the overall or specific subpopulation, (2) generating ideas for new hypotheses, (3) explaining uninterpretable findings between unadjusted and adjusted analyses, (4) guiding to present combined or separate models for each specific subpopulation, and (5) explaining heterogeneity in treatment effect. Often,

investigators present adjusted stratified results according to the presence or absence of an effect modifier. If the exposure interacts with multiple variables statistically or conceptually in the model, then the stratified findings (subgroup) according to each effect modifier may be presented. Otherwise, stratified analysis substantially reduces the power of the study due to the lower sample size in each stratum and may produce significant results by inflating type I error.[69] Therefore, a multivariable analysis involving an interaction term as opposed to a stratified analysis may be presented in the presence of an effect modifier.[70] Sometimes, a quantitative variable may emerge as a potential effect modifier for exposure and an outcome relationship. In such a situation, the quantitative variable should not be categorized unless a clinically meaningful threshold is not available in the study. In fact, the practice of categorizing quantitative variables should be avoided in the analysis unless a clinically meaningful cut-off is available or a hypothesis requires for it.[71] In an exploratory objective type, any possible interaction may be obtained in a study; however, the interpretation should be guided based on clinical implications. Similarly, some objective models may have more than one exposure or intervention and the association of each exposure according to the level of other exposure should be presented through adjusted analyses as suggested in the presence of interaction effects.[70]

Examples
A review of 428 articles from MEDLINE on the quality of reporting from statistical analyses of three (linear, logistic, and Cox) commonly used regression models reported that only 18.5% of the published articles provided interaction analyses,[17] even though interaction analyses can provide a lot of useful information.

Step 8: assessment of assumptions, specifically the distribution of outcome, linearity, multicollinearity, sparsity, and overfitting (reliability)
The assessment and reporting of model diagnostics are important in assessing the efficiency, validity, and usefulness of the model. Model diagnostics include satisfying model-specific assumptions and the assessment of sparsity, linearity, distribution of outcome, multicollinearity, and overfitting.[61 72] Model-specific assumptions such as normal residuals, heteroscedasticity and independence of errors in linear regression, proportionality in Cox regression, proportionality odds assumption in ordinal logistic regression, and distribution fit in other types of continuous and count models are required. In addition, sparsity should also be examined prior to selecting an appropriate model. Sparsity indicates many zero observations in the data set.[73] In the presence of sparsity, the effect size is difficult to interpret. Except for machine learning models, most of the parametric and semiparametric models require a linear relationship between independent variables and a functional form of an outcome. Linearity should be assessed using a multivariable polynomial in all model objectives.[62] Similarly, the appropriate choice of the distribution of outcome is required for model building in all study objective models. Multicollinearity assessment is also useful in all objective models.

Assessment of EVR in multivariable analysis can be used to avoid the overfitting issue of a multivariable model.[18]

Examples
Some review studies highlighted that 73.8%–92% of the articles published in MEDLINE had not assessed the model diagnostics of the multivariable regression models.[17 61 72] Contrary to the monotonically, linearly increasing relationship between systolic blood pressure (SBP) and mortality established using the Framingham's study,[74] Port *et al*[75] reported a non-linear relationship between SBP and all-cause mortality or cardiovascular deaths by reanalysis of the Framingham's study data set. This study identified a different threshold for treating hypertension, indicating the role of linearity assessment in multivariable models. Although a non-Gaussian distribution model may be required for modeling patient delay outcome data in cancer,[55] a study analyzed patient delay data using an ordinary linear regression model.[57] An investigation of the development of predictive models and their reporting in medical journals identified that 53% of the articles had fewer EVR than the recommended EVR, indicating over half of the published articles may have an overfitting model.[18] Another study[76] attempted to identify the anthropometric variables associated with non-insulin-dependent diabetes and found that none of the anthropometric variables were significant after adjusting for waist circumference, age, and sex, indicating the presence of collinearity. A study reported detailed sparse data problems in published studies and potential solutions.[73]

Step 9: report type of primary and sensitivity analyses (consistency)
Numerous considerations and assumptions are made throughout the research processes that require assessment, evaluation, and validation. Some assumptions, executions, and errors made at the beginning of the study data collection may not be fixable[13]; however, additional information collected during the study and data processing, including data distribution obtained at the end of the study, may facilitate additional considerations that need to be verified in the statistical analyses. Consistencies in the research findings via modifications in the outcome or exposure definition, study population, accounting for missing data, model-related assumptions, variables and their forms, and accounting for adherence to protocol in the models can be evaluated and reported in research studies using sensitivity analyses.[77] The purpose and type of supporting analyses need to be specified clearly in the statistical analyses to differentiate the main findings from the supporting findings. Sensitivity analyses are different from secondary or interim or subgroup analyses.[78] Data analyses for secondary outcomes are often referred to as secondary analyses, while data analyses of an ongoing study are called interim analyses and data analyses according to groups based on patient characteristics are known as subgroup analyses.

Examples
Almost all studies require some form of sensitivity analysis to validate the findings under different conditions. However, it is often underutilized in medical journals. Only 18%–20.3% of studies reported some forms of sensitivity

analyses.[77 78] A review of nutritional trials from high-quality journals reflected that 17% of the conclusions were reported inappropriately using findings from sensitivity analyses not based on the primary/main analyses.[77]

Step 10: provide methods for summarizing, displaying, and interpreting data (transparency and usability)

Data presentation includes data summary, data display, and data from statistical model analyses. The primary purpose of the data summary is to understand the distribution of outcome status and other characteristics in the total sample and by primary exposure status or outcome status. Column-wise data presentation should be preferred according to exposure status in all study designs, while row-wise data presentation for the outcome should be preferred in all study designs except for a case–control study.[24 32] Summary statistics should be used to provide maximum information on data distribution aligned with DGP and variable type. The purpose of results presentation primarily from regression analyses or statistical models is to convey results interpretation and implications of findings. The results should be presented according to the study objective type. Accordingly, the reporting of unadjusted and adjusted associations of each factor with the outcome may be preferred in the determinant objective model, while unadjusted and adjusted effects of primary exposure on the outcome may be preferred in the explanatory objective model. In prognostic models, the final predictive models may be presented in such a way that users can use models to predict an outcome. In the exploratory objective model, a final multivariable model should be reported with R^2 or area under the curve (AUC). In the association and interventional models, the assessment of internal validation is critically important through various sensitivity and validation analyses. A model with better fit indices (in terms of R^2 or AUC, Akaike information criterion, Bayesian information criterion, fit index, root mean square error) should be finalized and reported in the causal model objective study. In the predictive objective type, the model performance in terms of R^2 or AUC in training and validation data sets needs to be reported (figure 1).[20 21] There are multiple purposes of data display, including data distribution using bar diagram or histogram or frequency polygons or box plots, comparisons using cluster bar diagram or scatter dot plot or stacked bar diagram or Kaplan-Meier plot, correlation or model assessment using scatter plot or scatter matrix, clustering or pattern using heatmap or line plots, the effect of predictors with fitted models using marginsplot, and comparative evaluation of effect sizes from regression models using forest plot. Although the key purpose of data display is to highlight critical issues or findings in the study, data display should essentially follow DGP and variable types and should be user-friendly.[54 79] Data interpretation heavily relies on the effect size measure along with study design and specified hypotheses. Sometimes, variables require standardization for descriptive comparison of effect sizes among exposures or interpreting small effect size, or centralization for interpreting intercept or avoiding collinearity due to interaction terms, or transformation for achieving model-related assumptions.[80] Appropriate methods of data reporting and interpretation aligned with study design, study hypothesis,

and effect size measure should be specified in the statistical analysis section of research studies.

Examples

Published articles from reputed journals inappropriately summarized a categorized variable with mean and range,[81] summarized a highly skewed variable with mean and standard deviation,[57] and treated a categorized variable as a continuous variable in regression analyses.[82] Similarly, numerous examples from published studies reporting inappropriate graphical display or inappropriate interpretation of data not aligned with DGP or variable types are illustrated in a book published by Bland and Peacock.[83 84] A study used qualitative data on MRI but inappropriately presented with a Box-Whisker plot.[81] Another study reported unusually high OR for an association between high breast parenchymal enhancement and breast cancer in both premenopausal and postmenopausal women.[85] This reporting makes suspicious findings and may include sparse data bias.[86] A poor tabular presentation without proper scaling or standardization of a variable, missing CI for some variables, missing unit and sample size, and inconsistent reporting of decimal places could be easily noticed in table 4 of a published study.[29] Some published predictive models[87] do not report intercept or baseline survival estimates to use their predictive models in clinical use. Although a direct comparison of effect sizes obtained from the same model may be avoided if the units are different among variables,[35] a study had an objective to compare effect sizes across variables but the authors performed comparisons without standardization of variables or using statistical tests.[88]

A SAMPLE FOR WRITING STATISTICAL ANALYSIS SECTION IN MEDICAL JOURNALS/RESEARCH STUDIES

Our primary study objective type was to develop a (select from figure 1) model to assess the relationship of risk factors (list critical variables or exposures) with outcomes (specify type from continuous/discrete/count/binary/polytomous/time-to-event). To address this objective, we conducted a (select from figure 2 or any other) study design to test the hypotheses of (equality or superiority or non-inferiority or equivalence or futility) or develop prediction. Accordingly, the other variables were adjusted or considered as (specify role of variables from confounders, covariates, or predictors or independent variables) as reflected in the conceptual framework. In the unadjusted or preliminary analyses as per the (select from figure 3 or any other design features) DGP, (specify EBB preferred tests from online supplemental table 3 or any other appropriate tests) were used for (specify variables and types) in unadjusted analyses. According to the EBB practice for the outcome (specify type) and DGP of (select from figure 3 or any other), we used (select from online supplemental table 1 or specify a multivariable approach) as the primary model in the multivariable analysis. We used (select from figure 1) variable selection method in the multivariable analysis and explored the interaction effects between (specify variables). The model diagnostics including (list all applicable, including model-related assumptions, linearity, or multicollinearity or overfitting or distribution of outcome or sparsity) were also assessed using (specify appropriate methods) respectively. In such exploration, we

identified (specify diagnostic issues if any) and therefore the multivariable models were developed using (specify potential methods used to handle diagnostic issues). The other outcomes were analyzed with (list names of multivariable approaches with respective outcomes). All the models used the same procedure (or specify from figure 1) for variable selection, exploration of interaction effects, and model diagnostics using (specify statistical approaches) depending on the statistical models. As per the study design, hypothesis, and multivariable analysis, the results were summarized with effect size (select as appropriate or from figure 2) along with (specify 95%CI or other interval estimates) and considered statistically significant using (specify the side of p value or alternatives) at (specify the level of significance) due to (provide reasons for choosing a significance level). We presented unadjusted and/or adjusted estimates of primary outcome according to (list primary exposures or variables). Additional analyses were conducted for (specific reasons from step 9) using (specify methods) to validate findings obtained in the primary analyses. The data were summarized with (list summary measures and appropriate graphs from step 10), whereas the final multivariable model performance was summarized with (fit indices if applicable from step 10). We also used (list graphs) as appropriate with DGP (specify from figure 3) to present the critical findings or highlight (specify data issues) using (list graphs/ methods) in the study. The exposures or variables were used

in (specify the form of the variables) and therefore the effect or association of (list exposures or variables) on outcome should be interpreted in terms of changes in (specify interpretation unit) exposures/variables. List all other additional analyses if performed (with full details of all models in a supplementary file along with statistical codes if possible).

CONCLUDING REMARKS

We highlighted 10 essential steps to be reported in the statistical analysis section of any analytical study (figure 4). Adherence to minimum reporting of the steps specified in this report may enforce investigators to understand concepts and approach biostatisticians timely to apply these concepts in their study to improve the overall quality of methodological standards in grant proposals and research studies. The order of reporting information in statistical analyses specified in this report is not mandatory; however, clear reporting of analytical steps applicable to the specific study type should be mentioned somewhere in the manuscript. Since the entire approach of statistical analyses is dependent on the study objective type and EBB practice, proper execution and reporting of statistical models can be taught to the next generation of statisticians by the study objective type in statistical education courses. In fact, some disciplines (figure 5) are strictly aligned with specific study objective types. Bioinformaticians are oriented in studying

Figure 4 Summary of reporting steps, purpose, and evaluation measures in the statistical analysis section.
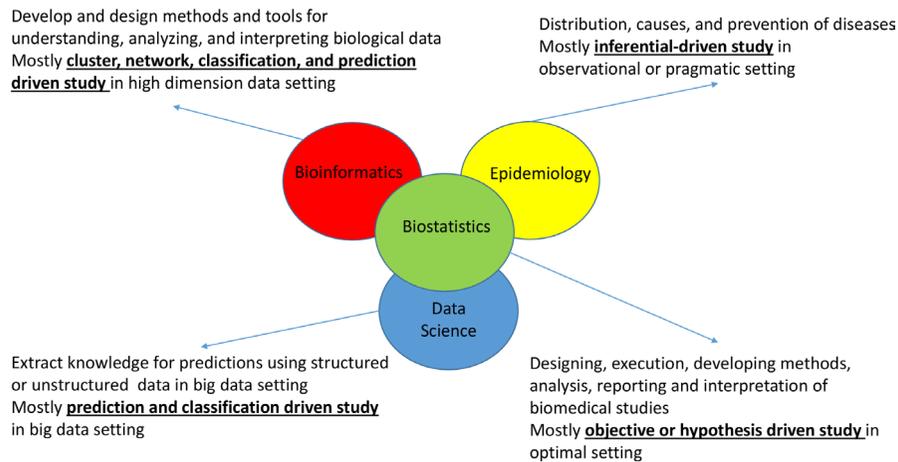
Develop and design methods and tools for understanding, analyzing, and interpreting biological data Mostly **cluster, network, classification, and prediction driven study** in high dimension data setting

Distribution, causes, and prevention of diseases Mostly **inferential-driven study** in observational or pragmatic setting

Extract knowledge for predictions using structured or unstructured data in big data setting Mostly **prediction and classification driven study** in big data setting

Designing, execution, developing methods, analysis, reporting and interpretation of biomedical studies Mostly **objective or hypothesis driven study** in optimal setting

Figure 5 Role of interrelated disciplines according to study objective type.

determinant and prognostic models toward precision medicine, while epidemiologists are oriented in studying association and causal models, particularly in population-based observational and pragmatic settings. Data scientists are heavily involved in prediction and classification models in personalized medicine. A common thing across disciplines is using biostatistical principles and computation tools to address any research question. Sometimes, one discipline expert does the part of others.[89] We strongly recommend using a team science approach that includes an epidemiologist, biostatistician, data scientist, and bioinformatician depending on the study objectives and needs. Clear reporting of data analyses as per the study objective type should be encouraged among all researchers to minimize heterogeneous practices and improve scientific quality and outcomes. In addition, we also encourage investigators to strictly follow transparent reporting and quality assessment guidelines according to the study design (https://www.equator-network.org/) to improve the overall quality of the study, accordingly STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) for observational studies, CONSORT (Consolidated Standards of Reporting Trials) for clinical trials, STARD (Standards for Reporting Diagnostic Accuracy Studies) for diagnostic studies, TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis OR Diagnosis) for prediction modeling, and ARRIVE (Animal Research: Reporting of In Vivo Experiments) for preclinical studies. The steps provided in this document for writing the statistical analysis section is essentially different from other guidance documents, including SAMBR.[5] SAMBR provides a guidance document for selecting evidence-based preferred methods of statistical analysis according to different study designs, while this report suggests the global reporting of essential information in the statistical analysis section according to study objective type. In this guidance report, our suggestion strictly pertains to the reporting of methods in the statistical analysis section and their implications on the interpretation of results. Our document does not provide guidance on the reporting of sample size or results or statistical analysis section for meta-analysis. The examples and reviews reported in this study may be used to emphasize the concepts and related implications in medical research.

ORCID iD
Alok Kumar Dwivedi http://orcid.org/0000-0003-4574-1761

REFERENCES
1 Oster RA, Devick KL, Thurston SW, *et al*. Learning gaps among statistical competencies for clinical and translational science learners. *J Clin Transl Sci* 2020;5:e12.
2 Sauerbrei W, Abrahamowicz M, Altman DG, *et al*. Strengthening analytical thinking for observational studies: the STRATOS initiative. *Stat Med* 2014;33:5413–32.
3 Thiese MS, Arnold ZC, Walker SD. The misuse and abuse of statistics in biomedical research. *Biochem Med* 2015;25:5–11.
4 Steven S, RMGTMK H. Statistical modeling methods: challenges and strategies. *Biostatistics & Epidemiology* 2020;4:105–39.
5 Dwivedi AK, Shukla R. Evidence-based statistical analysis and methods in biomedical research (SAMBR) checklists according to design features. *Cancer Rep* 2020;3:e1211.

6   Ocaña-Riola R. The use of statistics in health sciences: situation analysis and perspective. *Stat Biosci* 2016;8:204–19.

7   Wang MQ, Yan AF, Katz RV. Researcher requests for inappropriate analysis and reporting: a U.S. survey of consulting Biostatisticians. *Ann Intern Med* 2018;169:554–8.

8   Zhang G, Chen JJ. Biostatistics faculty and NIH awards at U.S. medical schools. *Am Stat* 2015;69:34–40.

9   Harrington D, D'Agostino RB, Gatsonis C, *et al*. New Guidelines for Statistical Reporting in the *Journal*. *N Engl J Med* 2019;381:285–6.

10   Nojima M, Tokunaga M, Nagamura F. Quantitative investigation of inappropriate regression model construction and the importance of medical statistics experts in observational medical research: a cross-sectional study. *BMJ Open* 2018;8:e021129.

11   Ciolino JD, Spino C, Ambrosius WT, *et al*. Guidance for biostatisticians on their essential contributions to clinical and translational research protocol review. *J Clin Transl Sci* 2021;5:e161.

12   Gosselin R-D. Insufficient transparency of statistical reporting in preclinical research: a scoping review. *Sci Rep* 2021;11:3335.

13   Brown AW, Kaiser KA, Allison DB. Issues with data and analyses: errors, underlying themes, and potential solutions. *Proc Natl Acad Sci U S A* 2018;115:2563–70.

14   Bacchetti P. Peer review of statistics in medical research: the other problem. *BMJ* 2002;324:1271–3.

15   Sima AP, Rodriguez VA, Bradbrook KE, *et al*. Incorporating professional recommendations into a graduate-level statistical consulting laboratory: a case study. *J Clin Transl Sci* 2020;5:e62.

16   Fernandes-Taylor S, Hyun JK, Reeder RN, *et al*. Common statistical and research design problems in manuscripts submitted to high-impact medical journals. *BMC Res Notes* 2011;4:304.

17   Real J, Forné C, Roso-Llorach A, *et al*. Quality reporting of multivariable regression models in observational studies: review of a representative sample of articles published in biomedical journals. *Medicine* 2016;95:e3653.

18   Bouwmeester W, Zuithoff NPA, Mallett S, *et al*. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012;9:e1001221–12.

19   Shmueli G. To explain or to predict? *Statistical Science* 2010;25:289–310.

20   Kent P, Cancelliere C, Boyle E, *et al*. A conceptual framework for prognostic research. *BMC Med Res Methodol* 2020;20:172.

21   Sainani KL. Explanatory versus predictive modeling. *Pm R* 2014;6:841–4.

22   Baser O. Choosing propensity score matching over regression adjustment for causal inference: when, why and how it makes sense. *J Med Econ* 2007;10:379–91.

23   Kent DM, Trikalinos TA, Hill MD. Are unadjusted analyses of clinical trials inappropriately biased toward the null? *Stroke* 2009;40:672–3.

24   Cummings P, Rivara FP. Reporting statistical information in medical Journal articles. *Arch Pediatr Adolesc Med* 2003;157:321–4.

25   Luo W, Phung D, Tran T, *et al*. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res* 2016;18:e323.

26   Spera G, Fresco R, Fung H, *et al*. Beta blockers and improved progression-free survival in patients with advanced HER2 negative breast cancer: a retrospective analysis of the ROSE/TRIO-012 study. *Ann Oncol* 2017;28:1836–41.

27   Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46:399–424.

28   Chen X, Zhou Y, Wang R, *et al*. Potential clinical value of multiparametric PET in the prediction of Alzheimer's disease progression. *PLoS One* 2016;11:e0154406.

29   Shaheen M, Echeverry D, Oblad MG, *et al*. Hepatitis C, metabolic syndrome, and inflammatory markers: results from the Third National Health and Nutrition Examination Survey [NHANES III]. *Diabetes Res Clin Pract* 2007;75:320–6.

30   Rajkumar P, Dwivedi AK, Dodoo CA, *et al*. The association between metabolic syndrome and hepatitis C virus infection in the United States. *Cancer Causes Control* 2020;31:569–81.

31   Knol MJ, Le Cessie S, Algra A, *et al*. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. *CMAJ* 2012;184:895–9.

32   Althouse AD, Raffa GM, Kormos RL. Your results, explained: clarity provided by row percentages versus column percentages. *Ann Thorac Surg* 2016;101:15–17.

33   Dwivedi AK, Mallawaarachchi I, Lee S, *et al*. Methods for estimating relative risk in studies of common binary outcomes. *J Appl Stat* 2014;41:484–500.

34   Cummings P. The relative merits of risk ratios and odds ratios. *Arch Pediatr Adolesc Med* 2009;163:438–45.

35   Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? *BMJ* 1998;316:989–91.

36   Oh S, Chung J, Baek S, *et al*. Postoperative expressive aphasia associated with intravenous midazolam administration: a 5-year retrospective case-control study. *J Int Med Res* 2020;48:030006052094875.

37   Chen ML, Gupta A, Chatterjee A, *et al*. Association between unruptured intracranial aneurysms and downstream stroke. *Stroke* 2018;49:2029–33.

38   Sturdik I, Krajcovicova A, Jalali Y, *et al*. Pathophysiology and risk factors for cholelithiasis in patients with Crohn's disease. *Physiol Res* 2019;68:S173–82.

39   Liao Y-T, Yang S-Y, Liu H-C, *et al*. Cardiac complications associated with short-term mortality in schizophrenia patients hospitalized for pneumonia: a nationwide case-control study. *PLoS One* 2013;8:e70142.

40   Doi SA, Furuya-Kanamori L, Xu C, *et al*. Controversy and debate: questionable utility of the relative risk in clinical research: paper 1: a call for change to practice. *J Clin Epidemiol* 2022;142:271–9.

41   Halsey LG. The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? *Biol Lett* 2019;15:20190174.

42   RLWNA L. The ASA statement on p-Values: context, process, and purpose. *The American Statistician*;70:129–33.

43   Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. *Int J Sports Phys Ther* 2014;9:726–36.

44   Weinberg CR. It's time to rehabilitate the P-value. *Epidemiology* 2001;12:288–90.

45   Ou F-S, Le-Rademacher JG, Ballman KV, *et al*. Guidelines for statistical reporting in medical journals. *J Thorac Oncol* 2020;15:1722–6.

46   Romano S, Goodman W, Tamura R, *et al*. Long-Term treatment of obsessive-compulsive disorder after an acute response: a comparison of fluoxetine versus placebo. *J Clin Psychopharmacol* 2001;21:46–52.

47   Wason JMS, Stecher L, Mander AP. Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? *Trials* 2014;15:364.

48   Tyler KM, Normand S-LT, Horton NJ. The use and abuse of multiple outcomes in randomized controlled depression trials. *Contemp Clin Trials* 2011;32:299–304.

49   Aiso I, Inoue H, Seiyama Y, *et al*. Compared with the intake of commercial vegetable juice, the intake of fresh fruit and komatsuna (Brassica rapa L. var. perviridis) juice mixture reduces serum cholesterol in middle-aged men: a randomized controlled pilot study. *Lipids Health Dis* 2014;13:102.

50   Allison DB, Antoine LH, George BJ. Incorrect statistical method in parallel-groups RCT led to unsubstantiated conclusions. *Lipids Health Dis* 2016;15:77.

51   Kim Y, Park S, Kim N-S, *et al*. Inappropriate survey design analysis of the Korean National health and nutrition examination survey may produce biased results. *J Prev Med Public Health* 2013;46:96–104.

52   Khera R, Angraal S, Couch T, *et al*. Adherence to methodological standards in research using the National inpatient sample. *JAMA* 2017;318:2011–8.

53   Weissgerber TL, Garcia-Valencia O, Garovic VD, *et al*. Why we need to report more than 'Data were Analyzed by t-tests or ANOVA'. *Elife* 2018;7. doi:10.7554/eLife.36163. [Epub ahead of print: 21 12 2018].

54   Weissgerber TL, Milic NM, Winham SJ, *et al*. Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol* 2015;13:e1002128.

55   Alok Kumar D, Nand DS, Suryanarayana D, *et al*. An epidemiological study on delay in treatment initiation of cancer patients. *Health* 2012;4. doi:10.4236/health.2012.42012

56   Dwivedi AK, Dwivedi SN, Deo S, *et al*. Statistical models for predicting number of involved nodes in breast cancer patients. *Health* 2010;2:641–51.

57   Poum A, Promthet S, Duffy SW, *et al*. Factors associated with delayed diagnosis of breast cancer in northeast Thailand. *J Epidemiol* 2014;24:102–8.

58   Ravdin PM, De Laurentiis M, Vendely T, *et al*. Prediction of axillary lymph node status in breast cancer patients by use of prognostic indicators. *J Natl Cancer Inst* 1994;86:1771–5.

59   Evans RG, Su D-F. Data presentation and the use of statistical tests in biomedical journals: can we reach a consensus? *Clin Exp Pharmacol Physiol* 2011;38:285–6.

60   Baker D, Lidster K, Sottomayor A, *et al*. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol* 2014;12:e1001756.

61   Zhang Y-Y, Zhou X-B, Wang Q-Z, *et al*. Quality of reporting of multivariable logistic regression models in Chinese clinical medical journals. *Medicine* 2017;96:e6972.

62   Sauerbrei W, Perperoglou A, Schmid M, *et al*. State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. *Diagn Progn Res* 2020;4:3.

63   Bursac Z, Gauss CH, Williams DK, *et al*. Purposeful selection of variables in logistic regression. *Source Code Biol Med* 2008;3:17.

64   Austin PC. Bootstrap model selection had similar performance for selecting authentic and noise variables compared to backward variable elimination: a simulation study. *J Clin Epidemiol* 2008;61:1009–17.

65 Chen R-C, Dewi C, Huang S-W, *et al*. Selecting critical features for data classification based on machine learning methods. *J Big Data* 2020;7:52.

66 Geleris J, Sun Y, Platt J, *et al*. Observational study of hydroxychloroquine in hospitalized patients with Covid-19. *N Engl J Med* 2020;382:2411–8.

67 Whitehouse AJO, Varcin KJ, Pillar S, *et al*. Effect of preemptive intervention on developmental outcomes among infants showing early signs of autism: a randomized clinical trial of outcomes to diagnosis. *JAMA Pediatr* 2021;175:e213298.

68 Ware RS. Reporting both Unadjusted and adjusted estimates is essential to the interpretation of randomized clinical trial results. *JAMA Pediatr* 2022;176:325–6.

69 Wang R, Ware JH. Detecting moderator effects using subgroup analyses. *Prev Sci* 2013;14:111–20.

70 Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. *Int J Epidemiol* 2012;41:514–20.

71 Figueiras A, Domenech-Massons JM, Cadarso C. Regression models: calculating the confidence interval of effects in the presence of interactions. *Stat Med* 1998;17:2099–105.

72 Ernst AF, Albers CJ. Regression assumptions in clinical psychology research practice-a systematic review of common misconceptions. *PeerJ* 2017;5:e3323.

73 Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ* 2016;352:i1981.

74 Li C, Chen Y, Zheng Q, *et al*. Relationship between systolic blood pressure and all-cause mortality: a prospective study in a cohort of Chinese adults. *BMC Public Health* 2018;18:107.

75 Port S, Demer L, Jennrich R, *et al*. Systolic blood pressure and mortality. *Lancet* 2000;355:175–80.

76 Wei M, Gaskill SP, Haffner SM, *et al*. Waist circumference as the best predictor of noninsulin dependent diabetes mellitus (NIDDM) compared to body mass index, waist/hip ratio and other anthropometric measurements in Mexican Americans--a 7-year prospective study. *Obes Res* 1997;5:16–23.

77 de Souza RJ, Eisen RB, Perera S, *et al*. Best (but oft-forgotten) practices: sensitivity analyses in randomized controlled trials. *Am J Clin Nutr* 2016;103:5–17.

78 Thabane L, Mbuagbaw L, Zhang S, *et al*. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol* 2013;13:92.

79 Kelleher C, Wagener T. Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software* 2011;26:822–7.

80 Althouse AD, Below JE, Claggett BL, *et al*. Recommendations for statistical reporting in cardiovascular medicine: a special report from the American heart association. *Circulation* 2021;144:e70–91.

81 DeLeo MJ, Domchek SM, Kontos D, *et al*. Breast MRI fibroglandular volume and parenchymal enhancement in BRCA1 and BRCA2 mutation carriers before and immediately after risk-reducing salpingo-oophorectomy. *AJR Am J Roentgenol* 2015;204:669–73.

82 Dontchos BN, Rahbar H, Partridge SC, *et al*. Are qualitative assessments of background parenchymal enhancement, amount of Fibroglandular tissue on Mr images, and mammographic density associated with breast cancer risk? *Radiology* 2015;276:371–80.

83 Bland M, Peacock J. *Statistical questions in evidence-based medicine*, 2000.

84 Chen JC, Cooper RJ, McMullen ME, *et al*. Graph quality in top medical journals. *Ann Emerg Med* 2017;69:453–61.

85 Telegrafo M, Rella L, Stabile Ianora AA, *et al*. Breast MRI background parenchymal enhancement (BPE) correlates with the risk of breast cancer. *Magn Reson Imaging* 2016;34:173–6.

86 Thompson CM, Mallawaarachchi I, Dwivedi DK, *et al*. The association of background parenchymal enhancement at breast MRI with breast cancer: a systematic review and meta-analysis. *Radiology* 2019;292:552–61.

87 Ramspek CL, Jager KJ, Dekker FW, *et al*. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2021;14:49–58.

88 Hansson O, Zetterberg H, Buchhave P, *et al*. Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *Lancet Neurol* 2006;5:228–34.

89 Goldstein ND, LeVasseur MT, McClure LA. On the convergence of epidemiology, biostatistics, and data science. *Harv Data Sci Rev* 2020;2. doi:10.1162/99608f 92.9f0215e6. [Epub ahead of print: 30 04 2020].