Article:
Mansournia, M.A., Collins, G.S., Nielsen, R.O. et al. (2021) A CHecklist for statistical Assessment of Medical Papers (the CHAMP statement): explanation and elaboration. British Journal of Sports Medicine, 55 (18). pp. 1009-1017. ISSN: 0306-3674

https://doi.org/10.1136/bjsports-2020-103652

A CHecklist for statistical Assessment of Medical Papers (The CHAMP Statement): Explanation and Elaboration

Mohammad Ali Mansournia*[1, 2], Rasmus O Nielsen[3, 4], Maryam Nazemipour*[5], Nicholas P Jewell[6, 7], Michael J Campbell[8], Douglas G Altman[9], Gary S Collins[9, 10]

1 Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

2 Sports Medicine Research Center, Neuroscience Institute, Tehran University of Medical Sciences, Tehran, Iran

3 Department of Public Health, Section for Sports Science, Aarhus University, Aarhus, Denmark

4 Research Unit for General Practice, Aarhus, Denmark

5 Psychosocial Health Research Institute, Iran University of Medical Sciences, Tehran, Iran

6 Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK

7 Division of Epidemiology & Biostatistics, School of Public Health, University of California, Berkeley, US

8 ScHARR, University of Sheffield, S1 4DA, Sheffield, UK

9 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK (deceased)

10 National Institute for Health Research Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

***Joint corresponding authors:**

Mohammad Ali Mansournia & Maryam Nazemipour

Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, PO Box: 14155-6446, Tehran, Iran. Tel: +98-21-88989123; Fax: +98-21-88989127; Email: mansournia_ma@yahoo.com , mansournia_ma@sina.tums.ac.ir

Psychosocial Health Research Institute, Iran University of Medical Sciences, Shahid Hemmat Highway, Tehran, P.O. Box:1449614535,IRAN, Email: nazemipour.m@iums.ac.ir; ma_nazemipour@yahoo.com

Tel: +(98-21) 86709; Fax: +(98-21) 88052248

Funding Statement

GSC was supported by the NIHR Biomedical Research Centre, Oxford, and Cancer Research UK (programme grant: C49297/A27294).

Abstract

Misuse of statistics in medical and sports science research is common and may lead to detrimental consequences on health care. Most authors, editors, and peer reviewers of medical papers will not have expert knowledge of statistics or may be unconvinced about the importance of applying correct statistics in medical research. Although there are guidelines on reporting statistics in medical papers, a checklist on the more general and commonly seen aspects of statistics to assess when peer-reviewing an article is needed. In this article, we propose a CHecklist for statistical Assessment of Medical Papers (CHAMP) comprising 30 items related to the design, conduct, data analysis, reporting and presentation, and interpretation of a research paper. Whilst CHAMP is primarily aimed at editors and peer reviewers during the statistical assessment of a medical paper, we believe it will serve as a useful reference to improve authors' and readers' practice in their use of statistics in medical research. We strongly encourage editors and peer reviewers (and readers) to consult CHAMP for assessing manuscripts for potential publication, and authors to ensure the validity of the general statistical aspects of their papers and reporting medical research.

Keywords: Statistics, Methodology, Medical research, Statistical checklist

The misuse of statistics by implementing flawed methodology in medical and sports science research can lead to unreliable or even incorrect conclusions. The consequences of flawed methodology can have undesirable consequences on public health, patient management and athlete performance[1]. Unfortunately, errors in the study design, statistical analysis, reporting, and interpretation of results are common in medical journals[2 3] and the quality of medical papers has been referred to as a scandal[4].

Sound methodology has been prioritized in the past decades, especially in high-impact factor journals. This is illustrated by the inclusion of more statistical editors and other methodologists (e.g. epidemiologists) in the review process. In addition, stakeholders in research have been encouraged to intensify their investments in statistical, epidemiological, and methodological education, such as training reviewers, providing online opportunities, developing (and extending) guidelines, and including methods content in regular scientific meetings[5]. There has also been a stronger emphasis on adherence to reporting guidelines (e.g., CONSORT, STROBE, STARD, REMARK and TRIPOD)[6-10].

Still, many medical and sports science journals do not involve statistical experts in the review process. This is unfortunate since the existence of basic statistical errors is more likely when authors, editors, and referees do not have sufficient knowledge of statistics and, worse, are unconvinced about the importance of correct statistics in medical research. Rarely do clinical journals systematically assess the use of statistics in submitted papers[11 12]. Thus, even after a paper is published in a scientific journal, it is necessary to read the content with some caution and pay careful attention to whether the statistical design and analysis were appropriate and the conclusions justified. Studies published in high-ranked journals are not immune from methodological or statistical flaws, which were not identified during the peer review process. Although some journals attempt to mitigate against such issues by using statisticians in the review process (as statistical reviewers or statistical editors), guidelines to assess methodological or statistical content in scientific papers would be useful when expert statistics reviewers are unavailable[5 13 14].

Whilst guidelines on how to report statistics in medical papers exist[15 16], we propose a general checklist to judge the statistical aspects of a manuscript for peer review. While it is clearly impossible to cover everything, we believe it would be useful to have a basic checklist for assessing the statistical methods used more broadly within medical and sports science research papers. Based on an extensive revision of a previous checklist[17], we describe a CHecklist for statistical Assessment of Medical Papers (CHAMP; Figure 1) comprising 30 items in the design, analysis, reporting and interpretation stages to aid the peer review of a submitted paper[18].

Explanation of the 30 items in the checklist

The 30 items in the checklist were selected based on a previous BMJ checklist[17], literature review, and experience of reviewing the statistical content of numerous papers submitted to a variety of

medical journals. The first author produced the checklist draft, the coauthors suggested addition or removal of the items, and all authors approved the final version. Other colleagues provided extensive comments on the paper and are listed in the Acknowledgments. Our checklist is not intended to, nor can it, cover all aspects of medical statistics. Our focus is rather on key issues that generally arise in clinical research studies. Therefore only important and common statistical issues encountered (including randomized controlled trials for which a separate checklist[17] has been suggested before) were included in the CHAMP. Using our checklist requires some primary knowledge of statistics, however, we provide a brief explanation for each item to shed light upon the item, and cite the relevant references for further details. The first six items relate to the design and conduct of research, whereas items 7 to 16 deal with data analysis, items 17 to 23 with reporting and presentation, and finally items 24 to 30 with interpretation.

ITEMS 1 – 7: DESIGN AND CONDUCT

Item 1: Clear description of the goal of research, study objective(s), study design, and study population

The research goal, study objectives, study design, and study and target populations must be clearly described so the editors of journals and readers can judge internal and external validity (generalizability) of the study.

Being explicit about the goal of research is a prerequisite for good science regardless of the scientific discipline. For such clarification, a 3-fold classification of the research goal may be used: 1) to describe; 2) to predict, which is equivalent to identifying "who" is at greater risk of experiencing the outcome; or 3) to draw a causal inference, which attempts to explain "why" the outcome occurs (e.g., investigating causal effects)[5 19].

The study objective refers to the rationale behind the study and points to the specific scientific question being addressed. For example, the objective of the HEx trial, a randomized controlled trial (RCT), was to evaluate the effect of heated water-based exercise training on 24-hour ambulatory blood pressure levels among resistant hypertensive patients[20].The study objective is usually provided in the introduction after the rationale has been established.

The study design refers to the type of the study, which is explained in the Methods section[21]. Examples of common study design include randomized controlled trials, cohort studies, case-control studies or cross-sectional studies[22]. The study design should be described in details. In particular, the randomization procedure in randomized controlled trials, follow-up time for cohort studies, control selection for case-control designs, and sampling procedure for cross-sectional studies should be adequately explained[6 7]. As a general principle, the study design must be explained sufficiently so that another investigator would be able to repeat the study exactly.

The study population refers to the source population from which data are collected whereas the target population refers to the population to whom we are going to generalize the study results; the relationship between these two populations may be characterized using inclusion and exclusion criteria and is crucial for assessing generalizability. Returning to the Hex trial, the study population was restricted to persons whose ages were between 40 and 65 years with resistant hypertension for more than 5 years[20]. For both trials and observational studies it is very important to know what proportion of the source population is studied, and what proportion of the intended data set features in the analysis data set. For example, the source population may be all patients admitted to a hospital with a certain condition over a certain period of time. However, the analysis data set may only be 50% of this, for various reasons such as patents refusing consent, measurements not taken, patients dropping out etc. For example in the HEX trial they had to screen 125 hypertensive patients to find 32 who met the inclusion criteria. This has some bearing on to whom heated-water based exercise training can be given and how likely it is to be relevant to other practitioners[20]

Item 2: Clear descriptions of outcomes, exposures/treatments and covariates, and their measurement methods

All variables considered for statistical analysis should be stated clearly in the paper, including outcomes, exposures/treatments, predictors and potential confounders/mediators/effect-measure modifiers. The measurement method and timing of measurement for each of these variables should also be specified. If the goal of the research is to draw a causal inference via observational studies, authors should also visualize their causal assumptions in a diagram[23 24]. To exemplify the concept, in an observational cohort study evaluating the effect of physical activity on functional performance and knee pain in patients with osteoarthritis[25], physical activity (exposure) was measured using the Physical Activity Scale for the Elderly, and functional performance and self-reported knee pain (outcomes) were measured by the Timed 20-meter Walk Test and the Western Ontario and McMaster Universities Osteoarthritis Index, respectively. Furthermore, depressive symptoms were considered as a potential confounder and measured using the Center for Epidemiologic Studies Depression Scale. All variables mentioned were measured at baseline and also in three annual follow-up visits[25].

Item 3: Validity of study design

The design should be valid i.e., it should match the research question and also should not introduce bias in the study results. For example, an editor should be able to assess whether the controls in a case-control study were adequately representative of the source population of the cases. Alternatively, in a clinical trial, it can be asked, first if there was one (or more) control groups and if so, if patients were randomized to treatment or control, and if so, whether the randomization method and allocation concealment were appropriate?

Item 4: Clear statement and justification of sample size

The manuscript should have a section clearly justifying the sample size[26]. When a sample size calculation is warranted, the sample size section should be described in enough detail to allow replication of the estimate, along with a clear rationale (supported by references) on choice of values used in the calculation, the outcome for which the calculation is based on, including the minimum clinically important effect size[27 28] . For example, typical sample size calculations aim to ensure the study enjoys a sufficiently large precision for estimates of occurrence measures such as risk or association measures like risk ratio[29 30], or that there is an adequate power to detect genuine effects if they exist (statistical tests). Attrition/lost-to-follow-up/non-response and design effects (e.g., due to clustering) should be taken into consideration. Some guidance for sample size calculation also exist in other areas, such as prediction model development and validation[31-33].

Item 5: Clear declaration of design violations and acceptability of the design violations

Design violations frequently occur in practice. Non-response in surveys, censoring (loss-to-follow-up or competing risks) in prospective studies[34], and non-compliance with the study treatments in randomized controlled trials should be declared explicitly in the paper[35]. Given validity of the design, the acceptability of violations should be assessed. For example, was an observed non-response/censoring proportion too high; what were the reasons for data loss, and is this level acceptable to achieve the scientific goals of the study?

Item 6: Consistency between the paper and its previously published protocol

The reviewer should identify inconsistencies with any published protocol (and where relevant, registry information) in terms of important features of the study including sample size, primary/secondary/exploratory outcomes, and statistical methods.

ITEMS 7 - 16: DATA ANALYSIS

Item 7: Correct and complete description of statistical methods

A separate part in the Methods section of the manuscript should be devoted to the description of the statistical procedures. Both descriptive and analytic statistical methods should sufficiently described so that the methods can be assessed by a statistical reviewer to judge their suitability and completeness in addressing the study objectives.

Item 8: Valid statistical methods used and assumptions outlined

The validity of statistical analyses relies on some assumptions. For example, the independent t-test for the comparison of two means requires three assumptions: independence of the observations, Normality, and homogeneity of variance. As another example, all expected values for a chi-square test must be more than 1, and at most 20% of the expected values can be less than 5. These statistical assumptions should be judged as a matter of context or assessed using appropriate methods such as a Normal probability plot for checking the Normality assumption[36]. In this regard, an alternative statistical test should be applied if some assumptions are clearly

violated. It should be noted that some statistical tests are robust against mild to moderate violations of some assumptions. For the t-test, lack of Normality and lack of homogeneity of variance do not necessarily invalidate the t-test, whereas lack of independence of the outcome variables will imply the results are invalid[37]. It has been demonstrated that independent t-test can be valid, but suboptimal for the Likert ordinal data even with a sample size of 20[38].

An important but often ignored aspect in practice is that ratio estimates like the estimated odds ratio, risk ratio and rate ratio are biased away from the null value. This bias is amplified with sparse data known as sparse-data bias[39]. A sign of sparse data is an unrealistically large ratio estimate or confidence limit which is simply an artifact of sparse data. For example, an OR>10 for a non-communicable disease should be considered as a warning sign for sparse-data bias. In the extreme, an empty cell leads to an absurd OR estimate of infinity, known as separation[40]. Special statistical methods such as penalization or Bayesian methods must be applied to decrease the sparse-data bias[40 41]. Some other important considerations in statistical analysis are:

i) accounting for correlation in the analysis of correlated data (e.g., variables with repeated measurements in longitudinal studies[42], cluster randomized trials[43], and complex surveys[44]);

ii)matching in the analysis of matched case-control and cohort data[45 46];

iii)ordering of several groups in the analysis;

iv) censoring in the analysis of survival data;

v) adjusting for baseline values of the outcome in the analysis of randomized clinical trials[27];

vi)correct calculation and interpretation of population attributable fraction[47 48];

vii) and adjusting for overfitting using shrinkage or penalization methods when developing a prediction model[49 50].

Item 9: Appropriate assessment of treatment effect or interaction between treatment and another covariate

Appropriate statistical tests should be used for the assessment of treatment effects and potential interactions. Assessment of overlapping treatment group-specific confidence intervals can be misleading[51-53]. Thus, the comparison of the confidence intervals of the treatment groups should not be used as an alternative to the statistical test of treatment effect. Moreover, comparing P-values for the treatment effect at each level of the covariate (e.g., men and women) should not be used as an alternative for an interaction test between the treatment and covariate. For example in the case of observing P-value<0.05 and P-value>0.05 in the levels one might incorrectly conclude that gender was a statistically significant predictor of outcome[54]. Similarly, we cannot conclude no effect modification if the confidence intervals of the subgroups are overlapping[55].

Item 10: Correct use of correlation and associational statistical testing

The misuse of correlation and associational statistical testing is not uncommon. As an example, correlation should not be used for assessing the agreement between two methods in methods-comparison studies[56]. To see why, two measures of X and Y are perfectly correlated but in poor agreement if X=2Y, but they are in poor agreement because X is twice Y. Likewise, we cannot infer that two methods agree well because the P-value is large enough using the statistical testing of the means such as paired t-test. In fact, a high variance of differences indicates poor agreement but also increases the chance that the paired t-test will result in a large P-value and thus the methods will appear to agree[1].

Item 11: Appropriate handling of continuous predictors

Reviewers should be wary of studies that have dichotomized or categorized continuous variables – this should be generally avoided.[57]. Bias, inefficiency and residual confounding may also result from dichotomizing/categorizing a continuous variable and using it as a categorical variable in a model. Continuous variables should be retained as continuous and their functional form be examined, as a linearity assumption may not be correct. Approaches for handling continuous predictors include fractional polynomials or regression splines[57-60].

Item 12: Confidence intervals do not include impossible values

A valid confidence interval should exclude impossible values. For instance, a simple Wald confidence interval for a proportion ($P \pm 1.96\sqrt{\frac{P(1-P)}{n}}$) is not valid when P is close to 0 or 1, and may yield negative values outside the possible range for a proportion ($0 \leq P \leq 1$)[61]. To remedy such conditions, the Wilson score or Agresti-Coull interval can be applied[6].

Item 13: Appropriate comparison of baseline characteristics between the study arms in randomized trials

In a randomized clinical trial, any baseline characteristic difference between groups should be due to chance (or unreported bias). Reviewers should look out for any statistical testing at the baseline as reporting P-values does not make sense[62]. The decision on which baseline characteristics (prognostic factors) are included in any adjustment should be pre-specified in the protocol and based on the subject-matter knowledge, not on P-values. The differences between groups in baseline characteristics should be identified by their size and discussed in terms of potential implications for the interpretation of the results

Item 14: Correct assessment and adjustment of confounding

An important goal of health research is drawing a causal inference. Here, the interest is in the causal effect of an exposure on the outcome. The major source of bias threatening causality studies, including observational studies as well as randomized studies (with small-to-moderate sample size), is confounding[63-65]. Confounding can be controlled in the design phase (e.g., through restriction or matching) or analysis phase (e.g., using regression models or propensity score

methods)[66 67]. Selection of confounders should be based on priori causal knowledge, often represented in the causal diagrams[23 68-70], *not* p-values (e.g., using stepwise approaches). Automated statistical procedures, such as stepwise regression, do not discriminate between confounders and other covariates like mediators or colliders which should not be adjusted for in the analysis. Moreover, stepwise regression is only based on the association between confounders and outcome, and disregards the association between the confounders and exposure. Thus, stepwise procedures should not be used for confounder selection. In practice, many confounders (and exposures and outcomes)[71 72] are time-varying, and the so-called "causal methods" should be applied for the appropriate adjustment of time-varying confounders[73]. Similarly, in studies evaluating the prognostic effect of a new variable, adjustment for existing prognostic factors should be routinely performed, and variable selection of the existing factors is not generally needed[49].

Item 15: On-support inference i.e., no model extrapolation to the region not supported by data

The goal of interest in many health studies is predicting an outcome from one or more explanatory variables using a regression model. The model is valid only within the range of observed data on the explanatory variables, and we cannot make prediction for people outside the range. This is known as model extrapolation[74]. Suppose we have found a linear relation between body mass index (BMI) and blood pressure (BP) based on the following equation:

$$BP = A + B * (BMI)$$

Now the intercept, A, cannot be interpreted since it corresponds to the blood pressure for a person with BMI of zero! The remedy is centering BMI and including the centered variable (BMI – average BMI) in the model so that the intercept refers to the blood pressure for a person with the average BMI in the population.

As another example, suppose the following linear relation holds in a randomized controlled trial:

$$BP = A + B * (TRT) + C * (BMI) + D * (TRT * BMI)$$

where TRT denotes treatment (1: intervention, 0: placebo) and TRT*BMI is the product term (interaction term) between treatment and BMI. In this model, the parameter B cannot be interpreted on its own because it is the mean difference in blood pressure between two treatment groups for a person with BMI of zero. Again, the solution is centering BMI and including centered BMI, and product term between TRT and centered BMI in the model so that the B' (coefficient of TRT in the new model) refers to the mean difference in blood pressure for a person with the average BMI in the population.

Item 16: Adequate handling of missing data

The methods used for handling missing data should be described and justified in relation to stated assumptions about the missing data (missing completely at random, missing at random, and missing not at random), and sensitivity analyses must be done if appropriate. Missing data[75] can introduce selection bias and should be handled using appropriate methods such as multiple imputation[76] and inverse probability weighting[77]. Naïve methods such as complete-case analysis, single imputation using the mean of the observed data, last observation carried forward, and the missing indicator method are statistically invalid in general and they can lead to serious bias[78].

ITEMS 17 - 23: REPORTING AND PRESENTATION

Item 17: Adequate and correct description of the data

, The mean and standard deviation provide a satisfactory summary of data for continuous variables that have reasonably a symmetric distribution. The standard error (SE) is not a sound choice to be used in place of SD[79]. A useful memory aid is to use standard Deviation to Describe Data and standard Error to Estimate parameters (Campbell MJ and Swinscow,TDV Statistics st Square One, London: BMJ Books 2009) Besides, "mean ± SD" is not suitable since it implies the range in which 68% of data are, not a relevant concept we are looking for, and "mean(SD)" should be reported instead[1]. In case of having highly skewed quantitative data, median and interquartile range (IQR) are more informative summary statistics for description. It should be noted that the mean/SD ratio<2 for positive variables is a sign of skewness[80]. Categorical data should be summarized as number and percentage[81]. For cohort data, a summary of follow-up time such as median and IQR should be reported.

Item 18: Descriptive results provided as occurrence measures with confidence intervals, and analytic results provided as association measures and confidence intervals along with P-values

The point estimates of the occurrence measures, for instance, prevalence, risk and incidence rate with 95% confidence intervals should be reported for descriptive objectives[81]. Alternatively, the point estimates of the association measures, for instance odds ratio, risk ratio and rate ratio with 95% confidence intervals along with P-values should be reported for analytic objectives as part of the results section[82].

Item 19: Confidence intervals provided for the contrast between groups rather than for each group

For analytic studies like randomized controlled trials, the 95% confidence intervals should be given for the contrast between groups rather than for each group[6]. For the blood pressure example mentioned above[20], the authors reported the mean of blood pressure with 95% confidence interval in each group but they should also have given themean difference in 24-hour ambulatory blood pressure levels between groups with 95% confidence interval since the aim of the trial was to compare treatment with control, not just report treatment and control outcomese.

Item 20: Avoiding selective reporting of analyses and P-hacking

All statistical analyses performed should be reported regardless of the results. P-hacking, playing with data to produce desired P-value (upward as well as downward), must be avoided[83-85]. This is probably difficult to assess as a reader/reviewer, but usually one would be clued in if there are many more analyses than those stated in the objectives or only statistically significant comparisons are presented when a larger pool of variables were identified in the methods.

Item 21: Appropriate and consistent numerical precisions for effect sizes, test statistics, and P-values, and reporting the P-values rather their range

P-values should be reported directly with one or two significant figures even if they are greater than 0.05, e.g., as P-value=0.09 or P-value=0.28. One should not focus on "statistical significance" or dichotomize P-values (eg $p<0.05$)[86-88] or express them as "0.000" or "NS". Nonetheless, spurious precision, too many decimals, in numerical presentation should be avoided[89 90]. For example, typically P-values less than 0.001 can be written as <0.001 without harm, and it does not make sense to present percentages with more than one decimal when the sample size is much less than 100.

Item 22: Providing sufficient numerical results that could be included in a subsequent meta-analysis

Meta-analyses of randomized trials and observational studies provide high levels of evidence in health research. Providing numerical results in individual studies contributing to subsequent meta-analysis is of special importance. Follow-up score and change score from the baseline are two possible approaches which can be applied to estimate treatment effect in randomized controlled trials[91]. While the follow-up score meta-analysis requires after-intervention mean and SD in two groups of intervention and placebo, the mean and SD of differences from the baseline are prerequisite for performing change-score meta-analysis. However, authors often only report mean and SD before and after intervention. The mean of the difference in each group can be calculated from the difference of the means, but calculating the SD of differences needs a guessed group-specific correlation between baseline and follow-up scores besides before- and after-intervention SD.

Item 23: Acceptable presentation of the figures and tables

Tables and figures are effective data presentation that should be properly managed[92-95]. Figures should be selected based on the type of variable(s) and appropriately scaled. The error bar graph as an illustration, can be used for displaying the mean and confidence interval. It is inappropriate to give s bar chart with a SE bar superimposed instead (a so called 'dynamite plunger plot[94]) . Tables should be able to stand on their own and include sufficient details such as labels, units, and values,

ITEMS 24 - 30: INTERPRETATION

Item 24: Interpreting the results based on association measures and 95% confidence intervals along with P-values, and correctly interpreting large P-values as indecisive results, not evidence of absence of an effect

The study results should be interpreted in light of the point estimate and appropriate association measures such as mean difference and 95% confidence interval as well as precise P-values. When testing a null hypothesis of no treatment effect, the P-value is the probability the statistical association would be as extreme as observed or more extreme, assuming that null hypothesis and all assumptions used for the test are correct. P-values for non-null effect sizes can also be computed. The point estimate is the effect size most compatible with the data in that it has P-value=1.00, while the 95% confidence interval shows the range of effect sizes reasonably compatible with the data in the sense of having P-value>0.05[87]. We should judge the clinical importance and statistical reliability of the results by examining both of the 95% confidence limits as well as looking at precise P-values, not just whether a P-value crosses a threshold or not[27 96]. It is incorrect to interpret P-value>0.05 as showing no treatment effect; instead it represents an ambiguous outcome[97 98]. It is not evidence that the effect is unimportant ("absence of evidence is not evidence of absence"); inferring unimportance requires that every effect size inside the confidence interval be considered unimportant[87].

Item 25: Using confidence intervals rather than post-hoc power analysis for interpreting the results of studies

Conceptually, it is not valid to interpret power as if it pertains to the observed study results[99-101]. Rather, power should be treated as part of study rationale and design before actual conduct begins, e.g., as in sample size calculations. Power does not correctly account for the observations that follow; for example, a study could have high power and observe a high P-value, yet still favor the alternative hypothesis over the null hypothesis[101]. The precision of results should be gauged using confidence intervals.

Item 26: Correctly interpreting occurrence or association measures

It will be crucial to interpret occurrence and association measures correctly. Odds ratios commonly provide examples of misinterpretation: if the event is rare, they can approximate risk ratios but they are not conceptually the same and will differ considerably if the event is common[102]. In a study with a risk of 60% in an exposed group and 40% in an unexposed group, the error in interpreting the odds ratio (2.25) as a risk ratio (1.5) is considerable. Prevalence in cross-sectional studies is another example, which sometimes has been incorrectly called 'risk'.

Item 27: Distinguishing causation from association and correlation

We should be cautious about the correct use of the technical terms such as effect, association and correlation. Association, meaning no independence, does not imply causation (and effect). Causal effect estimation requires measurement of exposure before outcome (temporality) as well as confounding adjustment. The correlation refers to a monotonic association between two variables. Therefore, no correlation does not imply no association.

Item 28: Results of pre-specified analyses are distinguished from the results of exploratory analyses in the interpretation

The results obtained from the pre-specified (a priori) analyses which have been already designed and mentioned in a protocol are much more reliable than the results obtained after data dredging (data-derived or post-hoc analyses).

Item 29: Appropriate discussion of the study methodological limitations

The methodological limitations of the study design and analysis should be discussed. Ideally, the probabilistic bias analysis, in which a probability distribution is assumed for the bias parameters, and bias is probabilistically accounted for using Monte-Carlo sensitivity or Bayesian analysis, should be performed for adjustment of uncontrolled confounding (e.g., due to an unmeasured confounder), selection bias (e.g., through missing outcome data) and measurement bias (e.g., subsequent to measurement error in the exposure)[103-105]. The authors should at least qualitatively discuss the main sources of bias and their impact on the study results[106 107].

Item 30: Drawing only conclusions supported by the statistical analysis and no generalization of the results to subjects outside the target population

The study interpretation must be based not only on the results but also in the light of the study population as well as any limitation in the design and analysis[74]. For example, if the study has been done in women, it cannot be necessarily generalized to the population of men and women.

Conclusion

The importance role of good statistic and sound methodology in medical research cannot be overstated. We strongly encourage authors to adhere to CHAMP for carrying out and reporting medical research, and to editors and reviewers for well-evaluating manuscripts for potential publication. We have only covered some basic items, and each type of study or statistical model (e.g., randomized trial, prediction model) has their own issues that ideally require statistical expertise. We appreciate that for some items in the checklist there is no unequivocal answer, and thus assessing the statistics of a paper may involve some subjectivity. Moreover, the questions in the checklist are not equally important e.g., papers with serious errors in design are statistically unacceptable regardless of how the data were analyzed, and aspects of presentations are clearly less important than other aspects. It is important to note that, statistical review, carried out by experienced statisticians, is the preferred way of reviewing statistics in research papers, more so

than what any checklist can achieve. We hope CHAMP improves authors' practice in their use of statistics in medical research and serve as a useful handy reference for editors and referees during the statistical assessment of medical papers.

Acknowledgments

We thank Sander Greenland, Stephen Senn, and Richard Riley for their valuable comments on an earlier draft of this paper.

Design and conduct				
1.	Clear description of the goal of research, study objective(s), study design, and study population	Yes	Unclear	No
2.	Clear descriptions of outcomes, exposures/treatments and covariates, and their measurement methods	Yes	Unclear	No
3.	Validity of study design	Yes	Unclear	No
4.	Clear statement and justification of sample size	Yes	Unclear	No
5.	Clear declaration of design violations and acceptability of the design violations	Yes	Unclear	No
6.	Consistency between the paper and its previously published protocol	Yes	Unclear	No
Data analysis				
7.	Correct and complete description of statistical methods	Yes	Unclear	No
8.	Valid statistical methods used and assumptions outlined	Yes	Unclear	No
9.	Appropriate assessment of treatment effect or interaction between treatment and another covariate	Yes	Unclear	No
10.	Correct use of correlation and associational statistical testing	Yes	Unclear	No
11.	Appropriate handling of continuous predictors	Yes	Unclear	No
12.	Confidence intervals do not include impossible values	Yes	Unclear	No
13.	Appropriate comparison of baseline characteristics between the study arms in randomized trials	Yes	Unclear	No
14.	Correct assessment and adjustment of confounding	Yes	Unclear	No
15.	On-support inference i.e., no model extrapolation to the region not supported by data	Yes	Unclear	No
16.	Adequate handling of missing data	Yes	Unclear	No
Reporting and presentation				
17.	Adequate and correct description of the data	Yes	Unclear	No
18.	Descriptive results provided as occurrence measures with confidence intervals, and analytic results provided as association measures and confidence intervals along with P-values	Yes	Unclear	No
19.	Confidence intervals provided for the contrast between groups rather than for each group	Yes	Unclear	No
20.	Avoiding selective reporting of analyses and P-hacking	Yes	Unclear	No

21.	Appropriate and consistent numerical precisions for effect sizes, test statistics, and P-values, and reporting the P-values rather their range	Yes	Unclear	No
22.	Providing sufficient numerical results that could be included in a subsequent meta-analysis	Yes	Unclear	No
23.	Acceptable presentation of the figures and tables	Yes	Unclear	No
Interpretation				
24.	Interpreting the results based on association measures and 95% confidence intervals along with P-values, and correctly interpreting large P-values as indecisive results, not evidence of absence of an effect	Yes	Unclear	No
25.	Using confidence intervals rather than post-hoc power analysis for interpreting the results of studies	Yes	Unclear	No
26.	Correctly interpreting occurrence or association measures	Yes	Unclear	No
27.	Distinguishing causation from association and correlation	Yes	Unclear	No
28.	Results of pre-specified analyses are distinguished from the results of exploratory analyses in the interpretation	Yes	Unclear	No
29.	Appropriate discussion of the study methodological limitations	Yes	Unclear	No
30.	Drawing only conclusions supported by the statistical analysis and no generalization of the results to subjects outside the target population	Yes	Unclear	No

Fig 1. Checklist for Statistical Assessment of Medical Papers

References

1. Altman DG. Practical statistics for medical research: CRC press 1990.
2. Thiese MS, Arnold ZC, Walker SD. The misuse and abuse of statistics in biomedical research. *Biochemia medica: Biochemia medica* 2015;25(1):5-11.
3. Thiese MS, Walker S, Lindsey J. Truths, lies, and statistics. *Journal of thoracic disease* 2017;9(10):4117.
4. Altman DG. The scandal of poor medical research: British Medical Journal Publishing Group, 1994.
5. Nielsen RO, Shrier I, Casals M, et al. Statement on methods in sport injury research from the 1st METHODS MATTER Meeting, Copenhagen, 2019. *Br J Sports Med* 2020 doi: 10.1136/bjsports-2019-101323 [published Online First: 2020/05/07]
6. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *International journal of surgery* 2012;10(1):28-55.
7. Vandenbroucke JP, Von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Annals of internal medicine* 2007;147(8):W-163-W-94.
8. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of internal medicine* 2003;138(1):W1-12.
9. Altman DG, McShane LM, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC medicine* 2012;10(1):51.
10. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine* 2015;162(1):W1-W73.
11. Altman DG. Statistical reviewing for medical journals. *Stat Med* 1998;17(23):2661-74. doi: 10.1002/(sici)1097-0258(19981215)17:23<2661::aid-sim33>3.0.co;2-b [published Online First: 1999/01/09]

12. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals. *Journal of general internal medicine* 1998;13(11):753-56.
13. Nielsen R, Shrier I, Casals M, et al. Statement on Methods in Sport Injury Research From the First METHODS MATTER Meeting, Copenhagen, 2019. *J Orthop Sports Phys Ther* 2020;50(5):226-33. doi: 10.2519/jospt.2020.9876 [published Online First: 2020/05/02]
14. Verhagen E, Stovitz SD, Mansournia MA, et al. BJSM educational editorials: methods matter. *Br J Sports Med* 2018;52(18):1159-60. doi: 10.1136/bjsports-2017-097998 [published Online First: 2017/08/19]
15. Lang T, Altman D. Basic statistical reporting for articles published in clinical medical journals: the SAMPL Guidelines In: Smart, Maisonneuve, Polderman A. *Handbook, European Association of Science Editors* 2013
16. Assel M, Sjoberg D, Elders A, et al. Guidelines for Reporting of Statistics for Clinical Research in Urology. *Eur Urol* 2019;75(3):358-67. doi: 10.1016/j.eururo.2018.12.014 [published Online First: 2018/12/21]
17. Gardner M, Machin D, Campbell M. Use of check lists in assessing the statistical content of medical studies. *Br Med J (Clin Res Ed)* 1986;292(6523):810-12.
18. Mansournia MA, Nielsen RO, Nazemipour M, et al. A CHecklist for statistical Assessment of Medical Papers: The CHAMP Statement. *British journal of sports medicine* 2020 (in press)
19. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. *Chance* 2019;32(1):42-49.
20. Guimaraes GV, de Barros Cruz LG, Fernandes-Silva MM, et al. Heated water-based exercise training reduces 24-hour ambulatory blood pressure levels in resistant hypertensive patients: a randomized controlled trial (HEx trial). *International journal of cardiology* 2014;172(2):434-41.
21. Centre for Evidence-Based Medicine. Study Designs 2020 [cited 2020 12 August]. Available from: https://www.cebm.net/2014/04/study-designs/.
22. Machin D, Campbell MJ. The design of studies for medical research: John Wiley & Sons 2005.
23. Etminan M, Collins GS, Mansournia MA. Using causal diagrams to improve the design and interpretation of medical research. *Chest* 2020;158(1):S21-S28.
24. Stovitz SD, Verhagen E, Shrier I. Distinguishing between causal and non-causal associations: implications for sports medicine clinicians. *British Journal of Sports Medicine* 2019;53(7):398-99. doi: 10.1136/bjsports-2017-098520
25. Mansournia MA, Danaei G, Forouzanfar MH, et al. Effect of physical activity on functional performance and knee pain in patients with osteoarthritis: analysis with marginal structural models. *Epidemiology* 2012:631-40.
26. Machin D, Campbell MJ, Tan SB, et al. Sample sizes for clinical, laboratory and epidemiology studies: John Wiley & Sons 2018.
27. Mansournia MA, Altman DG. Invited commentary: methodological issues in the design and analysis of randomised trials: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2018.
28. Cook JA, Julious SA, Sones W, et al. DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *bmj* 2018;363:k3750.
29. Bland JM. The tyranny of power: is there a better way to calculate sample size? *Bmj* 2009;339:b3985. doi: 10.1136/bmj.b3985 [published Online First: 2009/10/08]
30. Rothman KJ, Greenland S. Planning study size based on precision rather than power. *Epidemiology* 2018;29(5):599-603.

31. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *Bmj* 2020;368:m441. doi: 10.1136/bmj.m441 [published Online First: 2020/03/20]
32. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019;38(7):1276-96. doi: 10.1002/sim.7992 [published Online First: 2018/10/26]
33. Riley RD, Snell KIE, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: Part I - Continuous outcomes. *Stat Med* 2019;38(7):1262-75. doi: 10.1002/sim.7993 [published Online First: 2018/10/23]
34. Jungmalm J, Bertelsen ML, Nielsen RO. What proportion of athletes sustained an injury during a prospective study? Censored observations matter. *Br J Sports Med* 2020;54(2):70-71. doi: 10.1136/bjsports-2018-100440 [published Online First: 2019/05/23]
35. Nielsen RO, Bertelsen ML, Ramskov D, et al. Randomised controlled trials (RCTs) in sports injury research: authors-please report the compliance with the intervention. *Br J Sports Med* 2020;54(1):51-57. doi: 10.1136/bjsports-2019-100858 [published Online First: 2019/09/13]
36. Altman DG, Bland JM. Statistics notes: the normal distribution. *Bmj* 1995;310(6975):298.
37. Senn S. The t-test tool. *Significance* 2008;5(1):40-41.
38. Heeren T, D'Agostino R. Robustness of the two independent samples t-test when applied to ordinal scaled data. *Statistics in medicine* 1987;6(1):79-90.
39. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *bmj* 2016;352:i1981.
40. Mansournia MA, Geroldinger A, Greenland S, et al. Separation in logistic regression: causes, consequences, and control. *American journal of epidemiology* 2018;187(4):864-70.
41. Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Statistics in medicine* 2015;34(23):3133-43.
42. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis: John Wiley & Sons 2012.
43. Mansournia MA, Altman DG. Some methodological issues in the design and analysis of cluster randomised trials. *Br J Sports Med* 2019;53(9):573-75.
44. Korn EL, Graubard BI. Analysis of health surveys: John Wiley & Sons 2011.
45. Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. *International journal of epidemiology* 2013;42(3):860-69.
46. Mansournia MA, Jewell NP, Greenland S. Case–control matching: effects, misconceptions, and recommendations. *European journal of epidemiology* 2018;33(1):5-14.
47. Mansournia MA, Altman DG. Population attributable fraction. *Bmj* 2018;360:k757.
48. Khosravi A, Nielsen RO, Mansournia MA. Methods matter: population attributable fraction (PAF) in sport and exercise medicine. *British Journal of Sports Medicine* 2020
49. Riley RD, van der Windt D, Croft P, et al. Prognosis Research in Healthcare: concepts, methods, and impact: Oxford University Press 2019.
50. Steyerberg EW. Clinical prediction models: Springer 2019.
51. Bland JM, Peacock JL. Interpreting statistics with confidence. *The Obstetrician & Gynaecologist* 2002;4(3):176-80.
52. Austin PC, Hux JE. A brief note on overlapping confidence intervals. *Journal of Vascular Surgery* 2002;36(1):194-95. doi: https://doi.org/10.1067/mva.2002.125015
53. Mittal N, Bhandari M, Kumbhare D. A Tale of Confusion From Overlapping Confidence Intervals. *American Journal of Physical Medicine & Rehabilitation* 2019;98(1):81-83. doi: 10.1097/phm.0000000000001016
54. Matthews JN, Altman DG. Statistics Notes: Interaction 2: compare effect sizes not P values. *BmJ* 1996;313(7060):808.

55. Knol MJ, Pestman WR, Grobbee DE. The (mis)use of overlap of confidence intervals to assess effect modification. *European journal of epidemiology* 2011;26(4):253-54. doi: 10.1007/s10654-011-9563-8 [published Online First: 2011/03/19]
56. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-10. [published Online First: 1986/02/08]
57. Andersen PK, Skovgaard LT. Regression with linear predictors: Springer 2010.
58. Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables: John Wiley & Sons 2008.
59. Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer 2015.
60. Royston P, Sauerbrei W. Interaction of treatment with a continuous variable: simulation study of significance level for several methods of analysis. *Stat Med* 2013;32(22):3788-803. doi: 10.1002/sim.5813 [published Online First: 2013/04/13]
61. Mardani M, Rahnavardi M, Rajaeinejad M, et al. Crimean-Congo hemorrhagic fever among health care workers in Iran: a seroprevalence study in two endemic regions. *The American journal of tropical medicine and hygiene* 2007;76(3):443-45.
62. Senn S. Testing for baseline balance in clinical trials. *Statistics in medicine* 1994;13(17):1715-26.
63. Suzuki E, Tsuda T, Mitsuhashi T, et al. Errors in causal inference: an organizational schema for systematic error and random error. *Annals of epidemiology* 2016;26(11):788-93. e1.
64. Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. *European journal of epidemiology* 2015;30(10):1101-10.
65. Mansournia MA, Higgins JP, Sterne JA, et al. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology (Cambridge, Mass)* 2017;28(1):54.
66. Almasi-Hashiani A, Nedjat S, Mansournia MA. Causal Methods for Observational Research: A Primer. *Archives of Iranian Medicine (AIM)* 2018;21(4)
67. Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. *European heart journal* 2011;32(14):1704-08.
68. Nielsen RO, Bertelsen ML, Møller M, et al. Training load and structure-specific load: applications for sport injury causality and data analyses: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2018.
69. Nielsen RO, Bertelsen ML, Møller M, et al. Methods matter: exploring the 'too much, too soon'theory, part 1: causal questions in sports injury research. *British journal of sports medicine* 2020
70. Nielsen RO, Simonsen NS, Casals M, et al. Methods matter and the 'too much, too soon'theory (part 2): what is the goal of your sports injury research? Are you describing, predicting or drawing a causal inference?: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2020.
71. Nielsen RO, Bertelsen ML, Ramskov D, et al. Time-to-event analysis for sports injury research Part 1: time-varying exposures. *British journal of sports medicine* 2019;53(1):61-68.
72. Nielsen RO, Bertelsen ML, Ramskov D, et al. Time-to-event analysis for sports injury research part 2: time-varying outcomes. *British journal of sports medicine* 2019;53(1):70-78.
73. Mansournia MA, Etminan M, Danaei G, et al. Handling time varying confounding in observational research. *Bmj* 2017;359:j4587.
74. Altman DG, Bland JM. Generalisation and extrapolation. *Bmj* 1998;317(7155):409-10.
75. Altman DG, Bland JM. Missing data. *Bmj* 2007;334(7590):424-24.
76. Vickers AJ, Altman DG. Statistics notes: missing outcomes in randomised trials. *Bmj* 2013;346:f3438.
77. Mansournia MA, Altman DG. Inverse probability weighting. *Bmj* 2016;352:i189.

78. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj* 2009;338:b2393.
79. Altman DG, Bland JM. Standard deviations and standard errors. *Bmj* 2005;331(7521):903.
80. Altman DG, Bland JM. Detecting skewness from summary information. *British Medical Journal* 1996;313(7066):1200-01.
81. Nielsen RO, Debes-Kristensen K, Hulme A, et al. Are prevalence measures better than incidence measures in sports injury research?: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2019.
82. Nielsen RO, Bertelsen ML, Verhagen E, et al. When is a study result important for athletes, clinicians and team coaches/staff?: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2017.
83. Stovitz SD, Verhagen E, Shrier I. Misinterpretations of the 'p value': a brief primer for academic sports medicine: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2017.
84. Windt J, Nielsen RO, Zumbo BD. Picking the right tools for the job: opening up the statistical toolkit to build a compelling case in sport and exercise medicine research: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2019.
85. Nielsen RO, Chapman CM, Louis WR, et al. Seven sins when interpreting statistics in sports injury science: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2018.
86. McShane BB, Gal D, Gelman A, et al. Abandon statistical significance. *The American Statistician* 2019;73(sup1):235-45.
87. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology* 2016;31(4):337-50.
88. Rothman KJ, Greenland S, Lash TL. Precision and Statistics in Epidemiologic Studies. In: Rothman KJ, Greenland S, Lash TL, eds. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins 2008:148-67.
89. Altman DG, Bland JM. Statistics notes: Presentation of numerical data. *BMJ* 1996;312(7030):572.
90. Kordi R, Mansournia MA, Rostami M, et al. Troublesome decimals; a hidden problem in the sports medicine literature. *Scandinavian Journal of Medicine & Science in Sports* 2011;21(3):335-36. doi: 10.1111/j.1600-0838.2011.01312.x
91. Higgins J, Wells G. Cochrane handbook for systematic reviews of interventions. 2011
92. Schriger DL, Sinha R, Schroter S, et al. From submission to publication: a retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. *Ann Emerg Med* 2006;48(6):750-6, 56.e1-21. doi: 10.1016/j.annemergmed.2006.06.017 [published Online First: 2006/09/19]
93. Morris TP, Jarvis CI, Cragg W, et al. Proposals on Kaplan-Meier plots in medical research and a survey of stakeholder views: KMunicate. *BMJ Open* 2019;9(9):e030215. doi: 10.1136/bmjopen-2019-030215 [published Online First: 2019/10/03]
94. Freeman JV, Walters SJ, Campbell MJ. How to Display Data: Wiley 2009.
95. Vickers AJ, Assel MJ, Sjoberg DD, et al. Guidelines for Reporting of Figures and Tables for Clinical Research in Urology. *Eur Urol* 2020;78(1):97-109. doi: 10.1016/j.eururo.2020.04.048
96. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research: John Wiley & Sons 2008.
97. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature* 2019;567(7748):305-07. doi: 10.1038/d41586-019-00857-9 [published Online First: 2019/03/22]
98. Rafi Z, Greenland S. Semantic and cognitive tools to aid statistical science: Replace confidence and significance by compatibility and surprise. BMC Research Methodology, in press. *arXiv preprint arXiv:190908579[stat]* August 2020
99. Hoenig JM, Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician* 2001;55(1):19-24.

100. Bacchetti P. Peer review of statistics in medical research: the other problem. *Bmj* 2002;324(7348):1271-73.
101. Greenland S. Nonsignificance plus high power does not imply support for the null over the alternative. *Annals of epidemiology* 2012;22(5):364-68.
102. Janani L, Mansournia MA, Nourijeylani K, et al. Statistical issues in estimation of adjusted risk ratio in prospective studies. *Archives of Iranian medicine* 2015;18(10):0-0.
103. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data: Springer Science & Business Media 2011.
104. Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL, eds. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins 2008:345–80.
105. Lash TL, Fox MP, MacLehose RF, et al. Good practices for quantitative bias analysis. *Int J Epidemiol* 2014;43(6):1969-85. doi: 10.1093/ije/dyu149 [published Online First: 2014/08/01]
106. Altman DG, Bland JM. Uncertainty beyond sampling error. *Bmj* 2014;349:g7065.
107. Altman DG, Bland JM. Uncertainty and sampling error. *BMJ : British Medical Journal* 2014;349:g7064. doi: 10.1136/bmj.g7064