




Assessment of assumptions of statistical analysis methods in randomised clinical trials: the what and how

Anders Kehlet Nørskov ^{1,2,3}, Theis Lange,^{2,3}
Emil Eik Nielsen ^{1,4}, Christian Gluud,¹ Per Winkel,¹
Jan Beyersmann,⁵ Jacobo de Uña-Álvarez,⁶ Valter Torri,⁷
Laurent Billot,⁸ Hein Putter,⁹ Jørn Wetterslev ¹,
Lehana Thabane,¹⁰ Janus Christian Jakobsen^{1,4}

10.1136/bmjebm-2019-111268

For numbered affiliations see end of article.

Correspondence to:
Dr Anders Kehlet Nørskov,
Copenhagen Trial Unit, Centre
for Clinical Intervention
Research, Department
7812, Rigshospitalet,
Copenhagen 2100, Denmark;
anderskehlet@hotmail.com

Abstract

When analysing and presenting results of randomised clinical trials, trialists rarely report if or how underlying statistical assumptions were validated. To avoid data-driven biased trial results, it should be common practice to prospectively describe the assessments of underlying assumptions. In existing literature, there is no consensus on how trialists should assess and report underlying assumptions for the analyses of randomised clinical trials. With this study, we developed suggestions on how to test and validate underlying assumptions behind logistic regression, linear regression, and Cox regression when analysing results of randomised clinical trials.

Two investigators compiled an initial draft based on a review of the literature. Experienced statisticians and trialists from eight different research centres and trial units then participated in an anonymised consensus process, where we reached agreement on the suggestions presented in this paper.

This paper provides detailed suggestions on 1) which underlying statistical assumptions behind logistic regression, multiple linear regression and Cox regression each should be assessed; 2) how these underlying assumptions may be assessed; and 3) what to do if these assumptions are violated.

We believe that the validity of randomised clinical trial results will increase if our recommendations for assessing and dealing with violations of the underlying statistical assumptions are followed.

Introduction

The results of randomised clinical trials and systematic reviews hereof are, and should be, at the top of the hierarchy of evidence.^{1–5} However, several domains may bias trial results including the choice and conduct of statistical analyses.^{2,6–8}

In order to ensure validity of trial results and under some circumstances to optimise statistical power, most statistical methods do require checking and validation of underlying theoretical

assumptions.⁹ For example, the intervention effect estimated by analysis of covariance (ANCOVA) could under certain circumstances be erroneous if residuals are not normally distributed; and when using ANCOVA, power will often increase if data are log transformed.¹⁰ Likewise, if normality is violated and the sample size is small, then the Wilcoxon rank sum test can be three to four times more powerful than the independent samples t-test.⁹

Previously, we have undertaken a review of randomised clinical trials published in major medical journals in order to clarify whether the underlying assumptions behind the applied statistical tests were protocolised, assessed and reported.¹¹ The conclusion was that trialists rarely report if or how the underlying assumptions were validated.¹¹ Furthermore, the literature on the topic propose no clear recommendation on how to assess and report these underlying assumptions in relation to conducting a randomised clinical trial.¹¹

For decades, there have been a positive movement and ongoing strive to enhance transparency in medical research, for example, through the CONSORT statement and the EQUATOR network.^{6,12} However, it has not yet become common practice to prospectively report the methods used to assess the underlying assumptions of the statistical models used in the analysis of a randomised clinical trial.¹¹ It also seems essential to address what steps are to be taken if the underlying assumptions are not fulfilled and what criteria are used to decide whether an assumption is violated or not. By checking assumptions, we can strive to ensure that the appropriate statistical methodology is being used.

To facilitate clarity, completeness and transparency of reporting in randomised clinical trials, every step of the trial process, including assessments of assumptions underlying the chosen statistical methods, need to be thoroughly protocolised, described and reported.¹¹ The aim of this paper is: (1) to consider which underlying assumptions should be assessed when using logistic regression, linear regression and Cox regression during analysis of results from randomised clinical



© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Nørskov AK, Lange T, Nielsen EE, *et al.* *BMJ Evidence-Based Medicine* 2021;**26**:121–126.

trials; (2) to consider how to assess and validate these underlying assumptions and (3) how to deal with violations of these underlying assumptions.

We primarily focus on randomised clinical trials with two intervention arms, and logistic regression, linear regression and Cox regression as trialists commonly use these analysis methods in randomised clinical trials.¹¹ Furthermore, these regression analysis methods allow essential adjustments for stratification variables used in the randomisation as well as adjustments for baseline covariates.^{13 14}

Methods

Prior to beginning the development of the suggestions in this study, we performed a systematic survey of the literature to determine the current practice on the reporting of the assessment of the model assumptions for logistic, linear and Cox regression analyses in randomised trials. The results showed that the reporting of the results of model assumptions in trials is suboptimal.¹¹ The development of the recommendations in this study was done in two steps: (1) a systematic survey of the methodological literature to identify candidate assumptions for each method and (2) a consensus study among thirteen experts selected from academic clinical trials centres.

Systematic survey

The methodological literature was searched to identify candidate assumptions. Relevant databases were searched (PubMed, Cochrane Library, Google Scholar) using the search terms (assumption, statistical, analysis, randomi*) in February 2019. Based on the results of the systematic survey two investigators (TL and JCJ) developed a candidate list of assumptions and compiled an initial draft of the paper including:

- ▶ General considerations.
- ▶ Which assumptions to assess.
- ▶ How to assess if underlying assumptions are violated.

- ▶ Potential measures in case the assumptions are violated.

Consensus study of experts

The initial draft was distributed to invited selected investigators at different departments and institutions known in our network (see list of coauthors). We applied a Delphi-inspired process focusing on anonymised commenting for the investigators to be unbiased by opinions from other specific investigators. Each investigator at each institution was free to accept, reject, comment or suggest alternative methods preferably backed by arguments, results of empirical studies, results of simulation studies and other references to justify their comments and suggestions. All correspondence went exclusively and one-to-one through an independent facilitator (AKN). AKN collected all comments, assembled them into a report and wrote a compiled and anonymised summary of the comments. JCJ and TL then commented on the report from the facilitator and here on composed a revised draft of how to test for assumptions. The comments, the revised draft and the report from the facilitator were then sent to the external investigators for a next round of anonymised comments. This process was repeated six times until all coauthors could accept the final document.

Results

General considerations on which assumptions to assess, what methods to use and potential measures if the assumptions are not fulfilled

We strongly recommend that the protocol or the detailed statistical analysis plan for a trial should specify the key assumptions underpinning the analyses as well as how these assumptions should be assessed and what should be done if the assumptions are violated.¹¹

For all regression analyses, we recommend testing for major interactions between each covariate and the intervention variable. We recommend that the statistician, in turn, includes each possible first order interaction between included covariates and

| Logistic regression | Assumptions behind the analysis | How to assess if the assumptions are fulfilled | Potential measures if the assumptions are not fulfilled |
|---------------------|--|--|--|
| | A1: Observations are independent at least after adjusting for included covariates such as treatment site. | A1: Cannot be statistically established; use context knowledge. ^{21 22} | A1: In case of dependent data, methods appropriate to the analysis of clustered data should be considered (eg, mixed models, general estimating equations, robust methods and so on). |
| | A2: Linear effect on the log odds scale between continuous covariates and the outcome. | A2: Present Pearson residuals (observed values minus fitted values) against the fitted values of each continuous variable in a scatter plot. Patterns are taken as evidence against the assumption of a linear effect. ^{23 24} | A2: If the points are not close to a straight line, one option is to transform the continuous covariate, for example, to a factor variable (eg, into deciles or binary). ^{17 21 25} |
| | A3: No major interactions between each covariate and the intervention. ^{23 24} | A3: Include, in turn and one at a time, each possible first order interaction between included covariates and the intervention. For each combination, test if the interaction term is significant (a conservative approach is Bonferroni adjusted thresholds for significance, adjusted per number of possible interactions) and assess the effect size of the interaction term (the threshold must have been prespecified in the statistical analysis plan). The assumption behind logistic regression is violated if one or more of the tests of interaction are significant after multiplicity adjustment and the size is above a given threshold. ^{21 25} | A3: Significant and clinically important interactions justify a presentation of the results of the analysis both as an average across the subgroups, which show different treatment effects (ie, ignoring the interaction) and separately for each group. ^{21 25} |
| | A4: No relevant overdispersion. | A4: Since models in clinical trials are relatively simple one might just test if the deviance/df is significantly different from one or a prespecified distance from 1. | A4: Use a maximum likelihood estimate of the dispersion parameter in the analysis. |
| | A5: Sample size large enough to ensure valid CIs and p values based on asymptotic results (ie, all standard software). | A5: Check if the number of events is larger than 10 (rule of thumb) per independent parameter in the model. ²⁶ Unrealistically wide CIs of the independent parameters may also indicate that the number of events may be too low. | A5: Use bootstrap to compute CIs and p values. Alternatively (when the number of sites/stratification variables is high), use a mixed-effect model with a random-effects model for stratification variables. |

Table 2 Linear regression including ANCOVA

| Linear regression including ANCOVA | Assumptions behind the analysis | How to assess if the assumptions are fulfilled | Potential measures if the assumptions are not fulfilled |
|------------------------------------|---|---|---|
| | B1: Observations are independent at least after adjusting for included covariates such as treatment site. | B1: Cannot be statistically established; use context knowledge. ^{21 22} | B1: If the structure of dependence is known robust SEs can be used. In case of dependent data, methods appropriate to the analysis of clustered data should be considered (eg, mixed models) Otherwise the problem cannot be solved. |
| | B2: Normally distributed residuals. | B2: Do a quantile-quantile plot (known as a QQ plot) of the residuals (straight line required). ^{21 25} | B2 and B3: Transform outcome (typically log transformation or square root) or use an alternative generalized linear model that fits the distribution and/or use robust SEs. ^{21 25} Note that if the sample is large, say above 100 participants, deviations from normality will have very little effect, which in effect makes the assumption of normality less critical in typical randomised clinical trials. ²⁷ |
| | B3: Homogeneity of variances. | B3: Residuals plotted against covariates and fitted values. ^{21 24 25} | |
| | B4: Linear effect of the continuous covariate. | B4: Residuals plotted against the continuous covariate and fitted values. ^{21 25} | B4: If there is not a straight line one option is to transform the continuous variable. If a simple transformation does not work, one may transform the variable to a factor variable. ^{21 25} |
| | B5: No major interactions between each covariate and the intervention. ^{23 24} | B5: Include, in turn and one at the time, each possible first-order interaction between included covariates and the intervention. For each combination, test if the interaction term is significant (a conservative approach is Bonferroni adjusted thresholds for significance, adjusted per number of possible interactions) and assess the effect size of the interaction term (the threshold must have been prespecified in the statistical analysis plan). The assumption behind ANCOVA is violated if one or more of the tests of interaction are significant after multiplicity adjustment and the size is above a given threshold. ^{21 25} | B5: Significant and clinically important interactions justify a presentation of the results of the analysis both as an average across the subgroups, which show different treatment effects (ie, ignoring the interaction) and separately for each group. ^{21 25} If the covariate is continuous, a graphical display of the regression lines within each intervention group should be showed demonstrating the extent to which the lines are not parallel. The differences in all pair wise comparisons planned in the statistical analysis plan will now depend on the choice of observed covariate values used in the comparison. The range of the impact that this may have on the result of each comparison should be calculated. The primary analysis should still be the analysis without interactions unless the observed interaction is particularly large. In this case, the interpretation of the results may become impossible. ^{25 27} |

ANCOVA, analysis of covariance.

the intervention variable. For each combination, it should be assessed if the interaction term is significant and what the effect size is. Predefined threshold for significance should depend on the number of tests and although very restrictive Bonferroni adjusted thresholds may be considered (0.05 divided by number of possible interactions). Furthermore, it should be considered whether the interaction is assumed to have a clinically significant effect. If the interaction is concluded to be significant, it should be considered presenting separated analyses for each of the relevant variables (eg, for each site if there is significant interaction between the trial intervention and 'site') and an overall analysis including the interaction term in the model.

Visual inspection of different types of plots may be used to assess if underlying assumptions are violated (eg, visual inspection of histograms or plots of residuals). There are also several formal statistical tests, which may be used to test if certain underlying assumptions appear to be violated (eg, the Shapiro-Wilk test, Pearson's χ^2 test and the Anderson-Darling test).^{15 16} Visual inspection of plots has the limitation that it requires a subjective assessment of the plot in question and the subjective assessment may not be reliable and replicable.¹¹ The use of formal statistical tests also has limitations.¹¹ Formal tests for normality will, for example, often conclude that data are *not* normally distributed if

the data set is large, even when the departure from normality is inconsequential.¹¹ On the contrary, if the data set is small, serious departures from normality may not be detected by a formal test due to lack of power and the inherent asymmetry in formal hypothesis testing as discussed below.¹¹

We generally recommend using both graphical plots and formal statistical tests, and if there are discrepancies between these two assessments, then potential explanations for these discrepancies, and any action taken consequently, should be considered thoroughly and reported.¹¹

How to visually interpret statistical plots is a complex task and statistical knowledge and experience is needed.¹⁷ In the present paper, we will not consider how to interpret plots, but this is of course a prerequisite for a valid statistical assessment if plots are used.

To limit the potential bias, we suggest that any graphical and formal assessment of assumptions is performed blinded to the randomised interventions to ensure that the ensuing choice of method was not influenced by the effect of the intervention. We also encourage that all the plots used are published in a supplementary material as part of the main trial publication. That allows the reader to assess whether the methods are considered adequate. Furthermore, it should be documented how the choice of methods

Table 3 Survival analysis by Cox regression models

| Survival analysis by Cox regression models | Assumptions behind the analysis | How to assess if the assumptions are fulfilled | Potential measures if the assumptions are not fulfilled |
|--|--|---|---|
| | C1: Observations are independent at least after adjusting for included covariates such as treatment site. | C1: Cannot be statistically established; use context knowledge. ^{21 22} | C1: If the structure of dependence is known, robust SEs can be used. Otherwise the problem cannot be solved. |
| | C2: Independent censoring. | C2: Cannot be statistically established; use context knowledge. ^{21 22} | C2: If one suspects that dropout is informative, conduct a sensitivity analysis where the date of the occurrence of an event is set equal to the date of loss to follow-up. In other words, all patients lost to follow-up are assumed to experience the event on the date of censoring. In an alternative analysis, one should make the assumptions that these patients did not experience the event and the date of censoring is set equal to the date the trial was closed. Then, the range of bias caused by informative dropout can be assessed. |
| | C3: Proportional hazards. | C3: Proportional hazards between the compared intervention groups: -log log plot stratified by treatment and adjusted for the effects of all covariates (continuous and categorical) should show parallel lines. ^{21 22} | C3: Compute separate HR estimates for different proportions of follow-up, say first third, middle third and last third. If the estimates are not substantially different (corresponding to a clinically important difference), the range for this should be prespecified, the robustness of the Cox model ensures valid inference despite of the violation. If HRs differ substantially, they must all be reported along with a joint test for them being all equal to. ^{28 29} |
| | C4: Log linear effect of the continuous covariate. | C4: Martingale residuals plotted against time; there should be no time pattern. ³⁰ | C4: Transform the continuous variable to a factor variable (eg, into deciles). ^{21 22} |
| | C5: No major interactions between each covariate and the intervention. ^{21 23 25} | C5: Include, in turn and one at the time, each possible first order interaction between included covariates and the intervention. For each combination, test if the interaction term is significant (a conservative approach is Bonferroni adjusted thresholds for significance, adjusted per number of possible interactions) and assess the effect size of the interaction term (the threshold must have been prespecified in the statistical analysis plan). The assumption behind Cox regression is violated if one or more of the tests of interaction are significant after multiplicity adjustment and the size is above a given threshold. ^{21 25} | C5: See correspondent version of A5 and B5. Significant interactions justify a presentation of the results of the analysis adjusted for possible confounders and the stratification variables separate for groups defined by the covariate interacting with treatment (eg, if one centre in a multicentre trial shows a considerable different effect than the remaining centres). |
| | C6: Sample size large enough to ensure CIs and p values based on asymptotic results (ie, all standard software). | C6: Check if the number of events is larger than 10 (rule of thumb) per independent parameter in the model. ²⁵ Unrealistically wide CIs of the independent parameters may also indicate that the number of events may be too low. | C6: Use bootstrap to compute CIs and p values. Alternatively (when the number of covariates is high), use a mixed-effect model with a random-effect model for covariates. |

was arrived at. It is necessary that all assessments and statistical analyses are performed while the statistician remains blind to the randomised treatment allocation, that is, intervention groups should be coded as, for example, '1' and '2' and unnecessary variables that might compromise the blinding of the statistician should be excluded from the dataset used in the primary analyses.

Statistical reporting

Blinded data on all outcomes should preferably be analysed by two independent statisticians. Two independent statistical reports should be sent to the trial steering committee, and in case of discrepancies between the two statistical reports, possible reasons should be identified and consensus on the most correct result sought obtained. A final statistical report should then be prepared,

and all three statistical reports should be published as supplementary material.

Our detailed suggestions are presented in three tables, one for each analysis method, each presenting: 'Assumptions behind the analysis', 'How to assess if the assumptions are violated' and 'Potential measures if the assumptions are not fulfilled' (see tables 1–3).

Discussion

It can always be questioned whether it is possible and valid to present a 'cookbook' for statistical analyses. Nevertheless, it is difficult to assess the quality of a methodology or to improve it, if it is not described in detail. The present paper presents detailed suggestions on assessing and dealing with violation of underlying

assumptions for three frequently used statistical methods when analysing results of randomised clinical trials. Our recommendations are not exhaustive and following another methodology or systematic plan may also lead to valid results. This is particularly important to emphasise regarding the measures suggested for when the assumptions are *not* met, as several valid options often exist when violations of underlying assumptions occur. Nevertheless, the existing lack of transparency when trialists report how underlying statistical assumptions are assessed is striking—both in published protocols and in trial publications.¹¹ Other methodological aspects than tests for underlying assumptions have received thorough attention for decades.⁷ It is our intention that this paper should equally heighten attention on this important part of the methodology and be viewed as a supplement to existing recommendation.^{8 18}

When analysing results of randomised clinical trials, the primary objective will often be to clarify if a given intervention is beneficial or not. Other trial objectives might necessitate assessments of assumptions we have not included in our recommendations. As an example, we do plan to assess the proportional hazard assumption between the compared intervention groups. However, we do not recommend assessing the proportional hazards assumption for each covariate (eg, by visually inspect plots of Schoenfeld residuals both of continuous and categorical covariates¹⁹) because violations of this proportional hazard assumption will rarely influence the overall trial conclusions whether the intervention works or not. If the HRs for each covariate are of interest, then the proportional hazard assumption for this covariate should of course be assessed.

A potential limitation of the present paper is that we did not follow a strict Delphi methodology.²⁰ A Delphi process often includes several live meetings, SKYPE meetings and telephone conferences where recommendations are discussed, which has several advantages (in depth discussions, dedicated time for the project and so on). However, some researchers might, for example, be more persuasive than others or some researchers might have more charisma or be more famous than others, which might hinder recommending the optimal methodology if the famous researchers are not presenting the optimal viewpoints or arguments. Therefore, we chose to focus on anonymous commenting when considering our recommendations and our methodology is therefore described as a 'Delphi-inspired' methodology. Nevertheless, it may be a potential limitation that we did go through a full Delphi process.

Key messages

- ▶ The validity of results from randomised clinical trials is dependent on fulfilment of the underlying assumptions of the statistical analysis.
- ▶ There is no good consensus on how trialists should assess and report underlying assumptions for the analyses of randomised clinical trials.
- ▶ This paper presents suggestions on which underlying assumptions to test when using logistic regression, linear regression and Cox regression to analyse results of randomised clinical trials.
- ▶ The paper provides detailed suggestions on how these underlying assumptions may be assessed and what to do if the assumptions are violated.

Conclusion

This paper provides suggestions for assessing the validity of the assumptions underlying commonly used statistical analyses for randomised clinical trials, as well as suggestions for assessing and addressing violations. We believe that the validity of trial results will increase if our recommendations are followed.

Author affiliations

¹Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

²Department of Biostatistics, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

³Centre for Statistical Science, Peking University, Beijing, China

⁴Department of Cardiology, Holbæk Sygehus, Holbæk, Denmark

⁵Institute of Statistics, Universität Ulm, Ulm, Baden-Württemberg, Germany

⁶Department of Statistics and Operations Research & Biomedical Research Center (CINBIO), Universidade de Vigo, Vigo, Galicia, Spain

⁷IRCCS, Istituto Di Ricerche Farmacologiche Mario Negri, Milano, Italy

⁸The George Institute for Global Health, University of New South Wales, Sydney, New South Wales, Australia

⁹Leiden University Medical Center, Universiteit Leiden, Leiden, Zuid-Holland, The Netherlands

¹⁰Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

Correction notice This article has been corrected since it was published Online First. Minor formatting issues have been corrected, namely in the abstract.

Contributors JCJ and TL did the planning and initiated the project. All authors developed the framework and participated in writing the article. AKN is the guarantor of the article. All remaining authors are selected statistical professors, experts, methodologists and trialists from different parts of the world. The recommendations presented in this paper are derived from the authors' extensive experience and expertise in this field as well as from relevant and up to date literature. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement statement We did not involve patients or the public in our work.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs

Anders Kehlet Nørskov <http://orcid.org/0000-0002-6908-7832>

Emil Eik Nielsen <http://orcid.org/0000-0003-1108-9533>

Jørn Wetterslev <http://orcid.org/0000-0001-7778-1771>

References

- 1 Jakobsen J, Gluud C. The necessity of randomized clinical trials. *Br J Med Med Res* 2013;3:1453–68.
- 2 Jakobsen JC, Gluud C, Winkel P, *et al*. The thresholds for statistical and clinical significance - a five-step procedure for evaluation of intervention effects in randomised clinical trials. *BMC Med Res Methodol* 2014;14:34.
- 3 Jakobsen JC, Wetterslev J, Winkel P, *et al*. Thresholds for statistical and clinical significance in systematic reviews with meta-analytic methods. *BMC Med Res Methodol* 2014;14:120.

- 4 Piantadosi S. *Clinical trials: a methodologic perspective*. 2nd edn. Wiley, 2005.
- 5 Pocock SJ. Randomised clinical trials. *Br Med J* 1977;1:1661.
- 6 Equator N. Enhancing the quality and transparency of health research. Available: <http://www.equator-network.org/>
- 7 Garattini S, Jakobsen JC, Wetterslev J, et al. Evidence-Based clinical practice: overview of threats to the validity of evidence and how to minimise them. *Eur J Intern Med* 2016;32:13–21.
- 8 Schulz KF, Altman DG, Moher D, et al. Consort 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010;152:726–32.
- 9 Savović J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429–38.
- 10 Sawilowsky SS. Misconceptions leading to choosing the T test over the Wilcoxon Mann-Whitney test for shift in location parameter. *J Mod App Stat Meth* 2005;4:598–600.
- 11 Nielsen EE, Nørskov AK, Lange T, et al. Assessing assumptions for statistical analyses in randomised clinical trials. *BMJ Evid Based Med* 2019;24:185–9.
- 12 Consort. Available: <http://consort-statement.org/>
- 13 Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med* 2012;31:328–40.
- 14 Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. *BMJ* 2012;345:e5840.
- 15 Anderson TW, Darling DA. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann Math Statist* 1952;23:193–212.
- 16 Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika* 1965;52:591–611.
- 17 Hosmer DW, Hosmer T, Le Cessie S, et al. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997;16:965–80.
- 18 Chan A-W, Tetzlaff JM, Altman DG, et al. Spirit 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;158:200–7.
- 19 Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982;69:239–41.
- 20 Brown BB. *Delphi process: a methodology used for the elicitation of opinions of experts*. Santa Monica, CA: RAND Corporation, 1968.
- 21 Frydenberg SJ M. *An introduction to Stata for health researchers*. Fourth Edition. Stata Press, 2014.
- 22 Laville M, Segrestin B, Alligier M, et al. Evidence-based practice within nutrition: what are the barriers for improving the evidence and how can they be dealt with? *Trials* 2017;18:425.
- 23 Hosmer DW, Lemeshow S. *Applied logistic regression*. *Stat Med* 1989;10:1162–3.
- 24 Kragh Andersen P, Skovgaard LT. *Regression with linear predictors*. Springer, 2010.
- 25 Oliveira AG. *Biostatistics decoded*. Wiley, 2013.
- 26 Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and COX regression. *Am J Epidemiol* 2007;165:710–8.
- 27 LaVange LM, Durham TA, Koch GG. Randomization-based nonparametric methods for the analysis of multicentre trials. *Stat Methods Med Res* 2005;14:281–301.
- 28 Kleinbaum DG, Klein M. *Survival analysis - a self-learning text*. Third Edition. Springer, 2012.
- 29 Therneau TM, Grambsch PM. *Modeling survival data: extending the COX model*. Springer, 2000.
- 30 Klein JP, Moeschberger ML. *Survival analysis: techniques for censored and truncated data*. Springer, 1997.