

RESEARCH ARTICLE

Open Access



# A systematic review of the quality of conduct and reporting of survival analyses of tuberculosis outcomes in Africa

Moses M. Ngari<sup>1,2,3\*</sup>, Susanne Schmitz<sup>3</sup>, Christopher Maronga<sup>1,2</sup>, Lazarus K. Mramba<sup>4</sup> and Michel Vaillant<sup>3</sup>

## Abstract

**Background:** Survival analyses methods (SAMs) are central to analysing time-to-event outcomes. Appropriate application and reporting of such methods are important to ensure correct interpretation of the data. In this study, we systematically review the application and reporting of SAMs in studies of tuberculosis (TB) patients in Africa. It is the first review to assess the application and reporting of SAMs in this context.

**Methods:** Systematic review of studies involving TB patients from Africa published between January 2010 and April 2020 in English language. Studies were eligible if they reported use of SAMs. Application and reporting of SAMs were evaluated based on seven author-defined criteria.

**Results:** Seventy-six studies were included with patient numbers ranging from 56 to 182,890. Forty-three (57%) studies involved a statistician/epidemiologist. The number of published papers per year applying SAMs increased from two in 2010 to 18 in 2019 ( $P = 0.004$ ). Sample size estimation was not reported by 67 (88%) studies. A total of 22 (29%) studies did not report summary follow-up time. The survival function was commonly presented using Kaplan-Meier survival curves ( $n = 51$ , (67%) studies) and group comparisons were performed using log-rank tests ( $n = 44$ , (58%) studies). Sixty seven (91%), 3 (4.1%) and 4 (5.4%) studies reported Cox proportional hazard, competing risk and parametric survival regression models, respectively. A total of 37 (49%) studies had hierarchical clustering, of which 28 (76%) did not adjust for the clustering in the analysis. Reporting was adequate among 4.0, 1.3 and 6.6% studies for sample size estimation, plotting of survival curves and test of survival regression underlying assumptions, respectively. Forty-five (59%), 52 (68%) and 73 (96%) studies adequately reported comparison of survival curves, follow-up time and measures of effect, respectively.

**Conclusion:** The quality of reporting survival analyses remains inadequate despite its increasing application. Because similar reporting deficiencies may be common in other diseases in low- and middle-income countries, reporting guidelines, additional training, and more capacity building are needed along with more vigilance by reviewers and journal editors.

**Keywords:** Survival analysis, Time-to-event, Tuberculosis, Systematic review, Africa

\* Correspondence: [mngari@kemri-wellcome.org](mailto:mngari@kemri-wellcome.org)

<sup>1</sup>KEMRI/Wellcome Trust Research Programme, P.O Box 230, Kilifi 80108, Kenya

<sup>2</sup>The Childhood Acute Illness & Nutrition Network (CHAIN), Nairobi, Kenya

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Application of survival analyses, in this article referred to as '*Survival analyses methods (SAMs)*', have rapidly increased especially in oncology over the years [1]. They are used to analyze time-to-event outcomes and entail estimating; a) the probability of the outcome (event) of interest, b) the time the event occurs or c) exploring associations of time-to-event outcome with some independent predictors [2]. Therefore, SAMs usually provide more valuable information about how the probability of the event of interest changes with time compared to other standard statistical methods analyzing binary outcomes [2].

The probability of being event free at time  $t$ , usually denoted as *survival function* is commonly plotted using the Kaplan-Meier (KM) curve [2] while the probability of experiencing the event of interest; *the cumulative event function* is presented graphically using the Nelson-Aalen curve [3]. A life table is used to estimate and present survival time, but can only approximate the survival function within fixed intervals and is thus rarely used in survival analysis [4]. Log-rank tests are commonly used to compare the survival function between two or more groups [2].

The Cox Proportional Hazard (CPH) regression method, a semi-parametric model, is one of the most frequently used methods in survival regression analysis [5, 6]. The CPH model assumes the hazards are proportional over time (i.e. the hazard ratios are constant over time) [7]. Parametric proportional hazard models are similar but assume a specific statistical distribution for the hazard calculation and are considered more efficient because they estimate the baseline hazard rate [6, 8]. Additionally, SAMs have to take into account the non-informative censoring assumption (i.e. censoring time is statistically independent of their failure time) [2]. There are other broader considerations that are not SAMs specific and affect other applications of statistics such as appropriate assumptions when estimating sample size, lack of independence in presence of clustering or recurring events [2, 8–15].

Inappropriate conduct and low quality of reporting SAMs have been identified previously and may lead to incorrect conclusions [1, 16–18]. Previous published reviews of SAMs in medical research have found the quality of reporting SAMs inadequate [1, 16, 17, 19, 20]. The reviews included 764 studies (566 in oncology, 97 in cardiology, 73 in internal medicine, 14 in nephrology and 14 in acute lymphoblastic leukemia) conducted between 1991 and 2017. These reviews included only studies of non-communicable diseases predominately conducted in high income countries. All reviews identified significant deficiencies in reporting SAMs including non-reporting of sample size estimation and testing of the PH

assumption in the CPH regression models. In addition, there have been reports of inadequate and incomplete reporting of randomized trials and studies on infectious diseases without statisticians/epidemiologists in Africa [21, 22].

Tuberculosis (TB) is an infectious disease that requires treatment for at least six months. It is one of the leading causes of deaths from a single infectious agent globally and usually shows worse outcomes when it occurs among HIV infected patients [23]. Globally, the highest burden of TB is from Sub-Saharan Africa [24]. This article provides the first systematic review of the quality of reporting SAMs in studies of TB patients in Africa. In this study we aim to review the application and reporting of SAMs in studies of TB patients in Africa published from January 2010 to April 2020 in English.

## Methods

### Study design

We conducted a systematic review of studies from Africa that included TB patients and reported SAMs. TB end of treatment outcomes defined by World Health organization (WHO) formed the basis for the time-to-event analyses in this review (cured, completed treatment, failed treatment, died, defaulted, transferred out and successful treatment) [25]. The study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [26].

### Search strategy

A systematic search for eligible studies in MEDLINE via PubMed and EMBASE database was conducted in May 2020. The exact search terms are available in Additional file 1: Box 1.

### Selection criteria

Published papers were eligible for inclusion if they met the following criteria: i) the study population consisted of patients in Africa with TB (co-morbidity with other common infection like HIV was allowed); ii) follow-up data were available (i.e., cohort studies or randomized clinical trials); iii) SAM analysis methods were used; iv) the study was published between January 2010 and April 2020; v) the study was published in English language. Including papers published in the last ten years was deemed reasonable to capture recent trends in the application and reporting SAMs. We supplemented the search by reviewing references in the final list of articles that met eligibility criteria. Studies conducted in Africa but including sites outside Africa were excluded, however, where separate and complete analyses were conducted for each site, results from the African sites were included. We also excluded conference articles with

abstracts only, protocols, methodology papers, systematic reviews and meta-analyses.

### Screening of studies

The references from both databases were exported to Endnote X8 [27] where duplicates were removed. The remaining studies were exported into a screening software, Rayyan web app [28]. Study selection based on inclusion and exclusion criteria was conducted in a two-stage screening process: two assessors (MMN and CM) screened each reference first based on title and abstract and second based on the full text. All disagreements were resolved through discussion by the two assessors.

### Data extraction

We extracted data from the included studies in a data extraction template (Additional file 1: Appendix 1) designed in REDCap database [29]. The template was finalized following a piloting phase ensuring its suitability. Two authors (MMN and CM) independently performed the data extraction of each included reference; disagreements were resolved through discussion. In studies that performed more than one survival analysis, the main analysis was included. The following details were extracted: year of publication, publication journal, study design, involvement of a statistician/epidemiologist, collaboration with authors outside Africa, TB treatment outcome, number of study participants, reporting of follow-up time, graphic presentation of time-to-event, method used for group survival comparison (where applicable), type of survival regression models, method of testing underlying assumptions of any used regression models, statistical software used, reporting of sample size calculations, reporting exposure variables missing data, testing of interactions in regression models, reporting of lost-to-follow-up, censoring description and inclusion of multiple study sites/clusters. Information about involvement of a statistician/epidemiologist was extracted from the authors' affiliation, acknowledgement section or authors' information at the end of the manuscript and covered broad subject of statistics, biostatistics or epidemiology. Censoring description was assessed by checking studies that reported any mention of censoring, type of censoring, mention of non-informative censoring assumption and any method used when the non-informative censoring was violated or not assumed. Items covering broader statistical consideration like sample size estimation were included to help unravel the bigger methodological aspect, for example, a study with inadequate statistical power would yield non-conclusive results despite the SAMs used.

### Evaluation of quality of reporting survival analyses

The quality of reporting survival analyses was assessed using seven author-defined criteria (Table 1). Items included in the seven criteria were based on key elements of survival analyses identified by Altman et al [2, 6, 30, 31] and previous reviews [1, 16, 17] which were assessed by the authors, piloted and final items agreed upon. Through their experience, the authors, grouped the final items selected into the seven thematic survival analyses areas (the seven criteria). In brief, key survival analysis concepts and items previously evaluated were enumerated and organized into two domains: a) issues in design phase and b) statistical analysis phase. In design phase, sample size estimation and planned follow-up time were identified. Items identified during statistical analysis phase were grouped into five categories as presented in Table 1. Since this is a review of TB end-time outcomes, reporting consideration of recurring time-to-event in analysis was excluded.

### Statistical analysis

Frequency of studies reporting the seven evaluation criteria are reported with their respective percentage. We assessed the trend of the number of papers published across the years of publication from 2010 to 2019 using a Wilcoxon-type test for trend [32]. In a sub-analysis, we explored association of journal, year of publication and involvement of a statistician/epidemiologist with the quality of reporting (not reported, inadequate and adequately reported) using chi-square test/fisher's exact test. However, the results of the sub-analysis are only indication of possible associations as no power analysis was performed during study design. STATA/IC (version 15.1; StataCorp, College Station, TX, USA) was used to perform statistical analysis.

## Results

### Search results

The search yielded 1100 studies from MEDLINE (PubMed) and 1782 from EMBASE (Fig. 1). Six hundred and five duplicates were removed. We excluded 2177 studies after screening titles and abstracts. We reviewed the full text for 100 studies of which we excluded 24. Therefore, 76 studies were eligible for inclusion in the analysis. The full list of the 76 studies included is provided in the Additional file 1: Table S1.

### Characteristics of included studies

Characteristics of included studies are summarized in Table 2. Of the 76 studies, only one (1.3%) was a randomized trial, 54 (71%) were retrospective cohorts and 21 (28%) were prospective cohorts. Different time-to-event outcomes were evaluated with time to death ( $n = 72$ , (95%)) being the most common. The size of the

**Table 1** Criteria for evaluating quality of reporting SMAs

Criteria	Items assessed	Quality of reporting		
		Adequate	Inadequate	Not reported
Estimation of sample size	Statistical power; hypothesised effect estimate; effect size; alpha level; prevalence of exposure and probability of the expected outcome	All of these items reported for prospective studies. For retrospective studies, post hoc power estimation or detectable difference	At least one item was not reported	No sample size estimation information provided
Follow-up time	Start and exit dates and aggregate follow-up time (median/person-time)	Reported all these items	At least one item was not reported	None of the items was reported
Survival curves	Number of patients at risk at the bottom of the graph; markings to indicate when participants were censored; axes were clearly labelled and used different colors/type of lines to distinguish curves	Reported all these items	At least one item was not reported	No survival curve was plotted
Comparison of survival curves	Methods for group comparisons and their test results (p-values)	Reported all these items	At least one item was not reported	No comparison was done
Reporting measures of effect in SAMs	Measures of effect and uncertainty among studies reporting regression analysis	Correct measure of effect and uncertainty reported	Incorrect measure of effect or no measure of uncertainty	No measure of effect and uncertainty reported
Test of survival regression models underlying assumptions	Survival regression models used; statistical method used to test underlying assumptions and test results	Reported all these items	At least one of the items not reported	All the items not reported
Analysis of hierarchical clustering	Presence of clustering; methods of investigating heterogeneity and correct method for adjustment	Test of investigating heterogeneity and correct method for adjustment where there was evidence of heterogeneity reported	At least one of the items not reported	No consideration for clustering

studies ranged from 56 to 182,890 participants. Forty-three (57%) studies involved a statistician/epidemiologist in design or analysis. Collaborators from developed countries were included in 55 (72%) studies. STATA was the most commonly used software for data analysis in 40 (53%) studies, followed by SPSS (20%), SAS (12%) and R statistical programming ( $n = 7$ , (9.2%)). Five (6.6%) studies did not report the statistical software used [33–37]. Articles were most frequently published in PLOS One, International Journal of Tuberculosis and Lung Diseases (IJTLD) and BMC Infectious Disease; accounting for 53% of studies (Table 2). The number of published papers per year reporting SAMs increased from two in 2010 to 18 in 2019 ( $P = 0.004$ ) Fig. 2.

### Evaluation of reporting

#### Estimation of sample size

Very few ( $n = 9$ , (12%)) of the studies reported sample size estimation (Table 3), of which 3 (4.0%) did so adequately and 6 (7.9%) inadequately (Table 5).

#### Follow-up time

More than two thirds ( $n = 54$ , (71%)) of the studies reported follow-up time (Table 3): 52 (68%) adequately and 2 (2.6%) inadequately (Table 5).

#### Survival curves

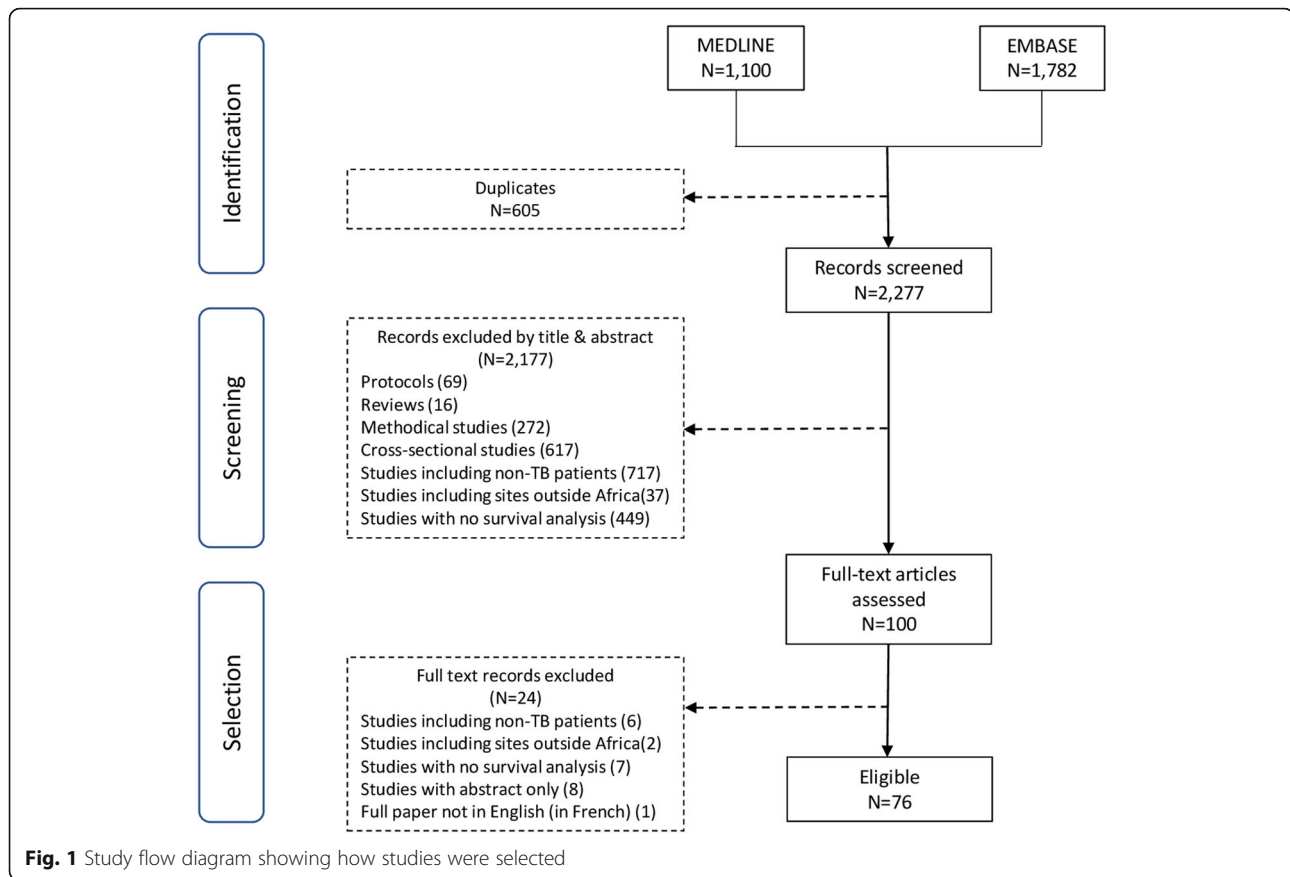
Survival curves were reported by 65 (86%) of the studies: Kaplan-Meier graphs were shown by 51 (67%) and Nelson-Aalen cumulative curves by 14 (18%) studies (Table 3). However, of the 14 studies reporting Nelson-Aalen cumulative curves, 9/14 (64%) were labelled as Kaplan-Meier [38–46]. Among the 65 studies reporting survival curves, 17/65 (26%) reported the number of patients at risk at each time point, 9/65 (14%) marked the survival time for the censored observations and all the 65 (100%) clearly labelled lines for different curves (Fig. 3). The reporting of survival curves was adequate in 1 (1.3%) and inadequate in 64 (84%) of the studies (Table 5).

#### Comparison of survival curves

The survival function estimator curves were compared between groups in 45 (59%) studies either by using log-rank test ( $n = 44$ , (58%)) or weighted log-rank test (Wilcoxon-Breslow-Gehan) ( $n = 1$ , (1.3%)) and all the 45 studies reported the test  $p$ -values (Table 3). All the 45 studies adequately compared the survival distributions (Table 5).

#### Reporting measures of effect

Seventy-four (97%) studies performed survival regression analysis: 67 (91%) using CPH model, 3 (4.1%) competing



risk analysis, and 4 (5.4%) parametric models. Two of the studies applying parametric proportional hazard models used Gompertz and Weibull probability distributions [47, 48], while 2 studies fitted an accelerated failure time parametric models, both using Weibull probability distributions [49, 50].

The two studies reporting parametric accelerated failure time [49, 50] and 69 studies performing Cox (67 studies) and parametric (2 studies) proportional hazard models reported time ratios (TR) and hazard ratios (HR) as the measure of effect respectively. Two of the three studies that performed competing risk analysis reported sub-distribution hazard ratios (SHR) [45, 51] while the other study reported HR [52]. All 74 studies reported 95% confidence intervals as measure of effect uncertainty (Table 3). The reporting of measures of effect was adequate among 73 (96%) and inadequate in 1 (1.3%) study (Table 5).

**Test of regression models underlying assumptions**

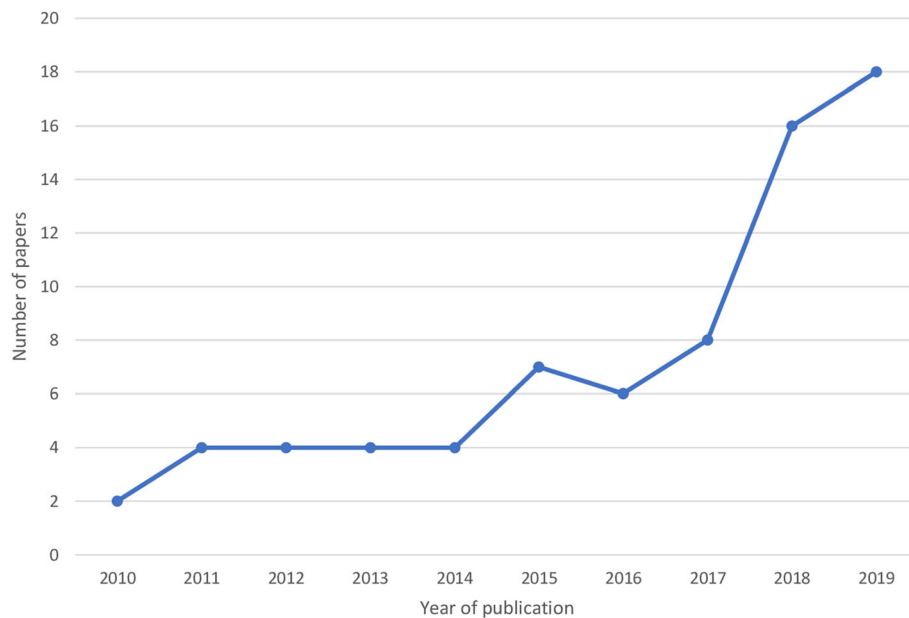
Among 67 studies that performed CPH regression analysis, 32/67 (48%) mentioned testing of the PH assumption (in the statistical methods section), however, only 2/67 (3.0%) reported the PH assumption test result [42, 53]. Where the PH assumption was violated, some

studies excluded individual predictors violating the assumption [54] or reported odds ratio rather than hazard ratios [55] or censored the analysis at 28 days for a study with follow-up of 12 weeks to meet the PH assumption [56]. Only one study [51] among the 3 that performed competing risk analysis mentioned testing the underlying PH assumption but did not report the test results. The two studies that used parametric PH methods tested the PH assumption using the Schoenfeld residual test and reported the results [47, 48]. All four studies (100%) that used parametric regression models reported testing the most fitting probability distribution using the maximum likelihood (LL), minimum Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC) and visual assessment of the Cox-Snell residual plots [47–50]. Three of the four studies (75%) reported the values of the LL, AIC and BIC for the different distributions assessed (Weibull, Exponential, Gompertz, log Logistic and Log-normal) and also plotted the Cox-Snell residual plots for all the distributions tested [47, 48, 50] (Table 3 and Fig. 3). The reporting of test of survival regression models’ assumption was adequate and inadequate among 5 (6.6%) and 32 (42%) studies respectively (Table 5).

**Table 2** Characteristics of studies included in the review

	All the studies (n = 76) n (%)
Study design	
Retrospective cohort	54 (71)
Prospective cohort	21 (28)
Randomized controlled trial	1 (1.3)
Type of study time-to-event outcome <sup>a</sup>	
Cured	9 (12)
Treatment complete	27 (36)
Treatment failure	16 (21)
Death	72 (95)
Default	37 (49)
Transfer out	5 (6.6)
Study size	
Median [IQR]	492 [286–1330]
Minimum to Maximum	56–182,890
Survival analyses objective	
Curve estimation	2 (2.6)
Survival regression analyses	16 (21)
Both	58 (76)
Involvement of statistician/epidemiologist	
Yes	43 (57)
Not reported	54 (71)
Authors affiliation <sup>b</sup>	
Country of focus only	21 (28)
Country of focus plus other African country	5 (6.6)
Country of focus plus developed country	55 (72)
Publication Journal	
PLOS One	22 (29)
International Journal of Tuberculosis and Lung Diseases (IJTLD)	9 (12)
BMC Infectious Disease	9 (12)
BMC Public Health	4 (5.3)
Clinical Infectious Disease	3 (3.9)
Others <sup>c</sup>	29 (38)
Statistical software used	
SPSS	15 (20)
SAS	9 (12)
STATA	40 (53)
R	7 (9.2)
Not reported	5 (6.6)

IQR interquartile range, <sup>a</sup>Studies evaluated more than one time-to-event outcome, <sup>b</sup>some studies had authors with African and developed countries affiliation and therefore the percentage > 100%, <sup>c</sup>AIDS-1, AIDS Respiratory Therapy-1, BMJ Thorax-1, American Journal of Tropical Medicine and Hygiene-1, Annals of Epidemiology-1, EclinicalMedicine-1, Infectious Diseases-1, Infection-1, Infectious Diseases of Poverty-2, International Journal of Infectious Diseases-2, International Journal of Mycobacteriology-2, International Journal of Pharmaceutical and Clinical Research-1, Journal of Acquired Immune Deficiency Syndromes-1, Lancet-1, Lancet Respiratory Medicine-1, PLOS Medicine-2, Pan African Medical Journal-1, The Pediatric Infectious Disease Journal –3, The Journal of Pediatrics-1, The Journal of Infectious Diseases-1, Tropical Medicine and International Health-2, Tropical Medicine and Health-1



**Fig. 2** Trend of the annual number of papers using SAMs from 2010 to 2019. Trend  $p$ -value = 0.004

#### **Analysis of hierarchical clustering**

Thirty seven (49%) studies had hierarchical clustering, some recruiting patients from multiple African countries [37, 57] or from one country but across widely dispersed hospitals with possible varying TB incidence. None of the 37 studies reported whether they assessed evidence of heterogeneity across the clusters. However, 9/37 (24%) of these studies reported consideration of clustering in the regression analysis (Table 4). All the 9 studies adequately controlled for the clustering in the analysis (Table 5).

#### **Description of other statistical methods**

Fifty (66%) studies reported censoring description. The majority ( $n = 46$ , (92%)) right censored participants following study completion, death, lost-to-follow-up or transfer out. Only one study reported investigating the non-informative censoring assumption by plotting observed survival times against values of the independent variables included in the regression models, and reported the assumption was not violated [35]. However, 4 (8%) studies reported considering the non-informative censoring assumption and adjusted the analysis using competing risk models (3 studies) and inverse probability censoring weighting (1 study) [58]. Four (5.2%) studies reported testing for some effect modification or interactions in the regression model [59–62] and provided stratified analyses where there was evidence of effect modification. A total of 70 (92%) of the studies did not report the proportion of missing exposure variables

data or how the missing data were handled in the analysis (Table 4).

#### **Overall evaluation**

Adequate reporting was high for reporting measures of effect and their uncertainty ( $n = 73$ , (96%)), follow-up time ( $n = 52$ , (68%)) and comparison of survival curves ( $n = 45$ , (59%)). However, adequate reporting was very low for sample size estimation ( $n = 3$ , (4.0%)), plotting of survival curves ( $n = 1$ , (1.3%)) and testing of underlying regression models assumptions ( $n = 5$ , (6.6%)). Approximately one quarter (24%) of studies adequately reported consideration of clustering in the regression models (Table 5).

In the sub-analyses, we found no evidence of journal, year of publication and involvement of a statistician/epidemiologist association with the quality of reporting SAMs (all  $P$ -values  $> 0.05$ ).

#### **Discussion**

In this systematic review of studies spanning over ten years, we found fundamental deficiencies in the reporting of survival analyses and an increasing trend in papers reporting SAMs annually. Sample size estimation, plotting of survival curves and assessment of regression underlying assumptions were rarely adequately reported. These deficiencies may lead to bias in reported measures of effect estimates and inaccurate conclusions. These are not isolated findings, as previous studies focusing on the quality of reporting SAMs [1, 16, 17, 19, 20], observational studies [63, 64] and even clinical trials [22, 65, 66]

**Table 3** Reporting of follow-up time, plotting of survival curves and survival regression analyses

	Number of articles n (%)
Sample size estimation reported	
Yes	9 (12)
Not reported	67 (88)
Reporting of follow-up time	
Median time	28 (37)
Person time	26 (34)
Not reported	22 (29)
Survival curves	
Kaplan-Meier	51 (67)
Nelson-Aalen	14 (18)
Not reported	11 (15)
Comparison of survival curves	
Log-rank test	44 (58)
Weighted log-rank test (Wilcoxon-Breslow-Gehan)	1 (1.3)
Not reported	31 (41)
Survival Regression models (N = 74)	
Cox proportional hazard	67 (91)
Competing risk analysis	3 (4.1)
Parametric proportional hazard	2 (2.7)
Parametric accelerated failure time	2 (2.7)
Reported regression models assumptions tested <sup>a</sup>	
Cox PH (N = 67)	
Visual (graphical log-log plots)	11 (16)
Schoenfeld residuals test	21 (31)
Not reported	35 (52)
Competing risk analysis (N = 3)	
Schofield residuals test	1 (33)
Not reported	2 (67)
Parametric Methods (N = 4)	
Information theory (LL, AIC, BIC)	4 (100)

<sup>a</sup>These are reported methods used to test the underlying assumptions in the statistical methods section and not the actual number of studies that reported the test results, PH-Proportional Hazard, LL-likelihood values, AIC-Akaike Information Criteria, BIC-Bayesian Information Criteria

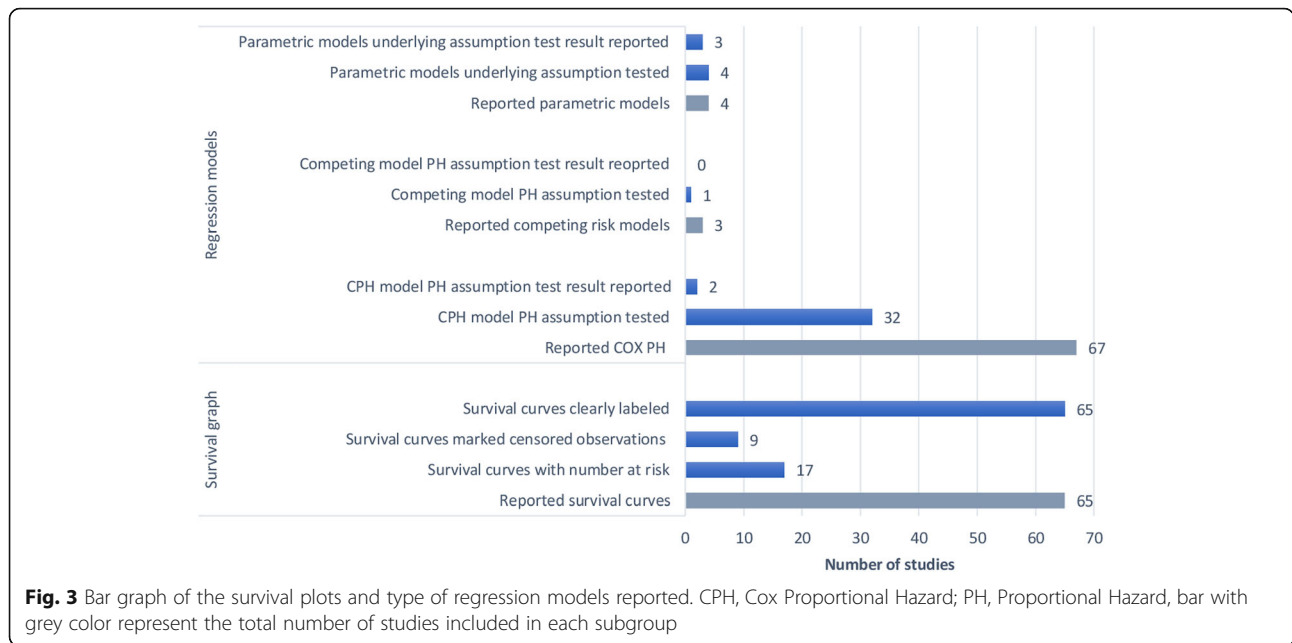
reported similar inadequacy. However, our analysis showed adequate reporting of effect measures.

Unlike a previous review of studies published in cancer journals, follow-up time and comparison of survival curves were frequently reported adequately [16]. However, some authors did not correctly distinguish Kaplan-Meier and Nelson-Aalen cumulative curves. Although Kaplan-Meier curves were commonly reported, in practice Nelson-Aalen curves plotting cumulative proportions of patients who experience the event are more informative [67]. In two previous reviews [1, 16] and this review, log-rank test was frequently reported probably because of its simplicity [2]. However, its *p*-value may not provide much information about the probabilities of

an event at different time points and therefore providing a measure of survival time in each group like median survival time would be more useful. A log-rank test is most appropriate when the PH assumption is met [68–70], an alternative is the weighted log-rank test which assigns weights proportional to the contribution of each failure time [68, 71–74] but was rarely used.

Just like previous reviews [1, 17, 19], CPH regression models were used in the majority studies. Although 48% of the studies did mention that they evaluated the PH assumption using either visual (graphical log-log plots) or residuals tests (Schoenfeld), only 3.0% reported the test results. In a review of 14 studies that used CPH models, none reported assessing the PH assumption [1]





**Fig. 3** Bar graph of the survival plots and type of regression models reported. CPH, Cox Proportional Hazard; PH, Proportional Hazard, bar with grey color represent the total number of studies included in each subgroup

while another review of application of SAMs in clinical trials found only 2/28 (7.1%) reported assessing the PH assumptions [19]. Similarly, amongst 112 Chinese Oncology studies that used CPH models, none reported assessing the PH assumption [20]. Only four studies used parametric methods and reported assessment of the underlying assumptions. All the four studies involved a statistician/epidemiologist, a demonstration of the central role they play. When correctly specified, parametric

models are more efficient and informative because they provide an estimate of baseline hazard ratio that can be used in predicting absolute risks [30, 75].

**Table 4** Reporting of important other analytic considerations

	Number of articles n (%)
Analysis of studies with hierarchical clustering (N = 37)	
No clustering consideration	28 (76)
Variance corrected method	4 (11)
Frailty models	3 (8.1)
Multilevel regression	2 (5.4)
Censoring description	
Yes	50 (66)
No	26 (34)
Test of effect modification/interaction (N = 4)	
Likelihood ratio test	2 (50)
Chi-square test of homogeneity	1 (25)
Not reported	1 (25)
Handling of missing exposure data	
Single imputation	2 (2.6)
Multiple imputation	3 (4.0)
A missing category included	1 (1.3)
Not reported	70 (92)

Our findings suggest many authors were not aware of the alternatives to use when PH assumption is violated and resulted to incorrect methods like excluding independent variables found to have violated the PH assumption [54]. When the PH assumption is violated for some continuous variables, creating binary or ordinal variables could be an option [30]. Alternatively, the variables could be included as time-varying predictors or time stratified analysis could be performed [30]. Parametric accelerated failure time models measure the effect of the covariate on a time scale rather than hazard scale and do not assume the PH assumption. They have been shown to be more robust in oncology and may be considered too [9, 76]. Restricted mean survival time (RMST) which reports the difference in RMST as a measure of effect at suitable follow-up time as been suggested as other alternative when PH assumption is violated [77]. A possible reason for many authors to not report test results of model assumptions may be journal's restrictions in the number of tables/figures allowed. The three studies that extensively reported the AIC, LL and BIC test results and plotted the cox-Snell plots were published in journals that do not limit number of tables/figures [47–49]. However, in the sub-analysis we found no evidence of association between the journal of publication and any of the reporting criteria. We would recommend journals to encourage authors to report these test results in the supplementary appendix as an extension of the statistical methods.

**Table 5** Overall quality of reporting SAMs

Criteria	Quality of reporting		
	Not reported	Inadequate	Adequate
Sample size estimation	67 (88)	6 (7.9)	3 (4.0)
Follow-up time	22 (29)	2 (2.6)	52 (68)
Plotting of survival curves	11 (14)	64 (84)	1 (1.3)
Comparison of survival curves	31 (41)	0	45 (59)
Test of regression underlying assumptions	39 (51)	32 (42)	5 (6.6)
Reporting measures of effect	2 (2.6)	1 (1.3)	73 (96)
Analysis of hierarchical clustering (N = 37)	28 (76)	0	9 (24)

In presence of competing events, the Kaplan-Meier function produces biased estimates. When the time-to-event of interest is treatment success, it is plausible to assume other treatment outcomes such as 'death', 'lost-to-follow-up' and 'transfer out' were informative censored and thus considered as competing events. Fine and Gray non-parametric test comparing the cumulative incidence functions without requirement of non-informative censoring could be used in such settings [10, 78–81]. However, application of this method was rare in this review and one of the studies using the method, incorrectly reported hazard ratios rather than sub-distribution hazard ratios [52]. Other methods like inverse probability censoring weighting and some

proposed methods using predicted long-term vital status may yield more accurate measures of effect estimates [82, 83]. Violation of non-informative censoring assumption may result in biased measure of effect estimates and thus should be investigated and appropriate adjustment made in the analysis although this was rarely done in the papers reviewed [82, 83].

Only 4% studies adequately reported the estimation of sample size, which is a key ingredient in any study design and a factor in determining the power to yield valid results. In a systematic review of lymphoblastic leukemia literature, 4/14 (29%) studies reported estimation of the study size which is slightly higher than our finding [1]. Since 71% of the studies were retrospective cohorts in

**Table 6** Recommendation for reporting survival analyses methods

Section	Recommendation
Study design	<ul style="list-style-type: none"> <li>Define the study time-to-event outcome.</li> <li>Report the sample size and sample estimation methods providing all the assumptions made in calculating sample size.</li> <li>Report the planned fixed length of follow-up (days, months, years).</li> </ul>
Statistical methods and results	<ul style="list-style-type: none"> <li>Report beginning and end dates of each event under observation.</li> <li>Report the total time under observation using standard epidemiological units like person-years and median time. Where the aim is to compare groups of participants, in addition to total time observed, report total time and median time stratified by the groups.</li> <li>Report number of participants lost-to-follow-up, how censoring was done and if non-informative censoring assumption was evaluated.</li> <li>Report total number of time-to-event outcome events observed, and events per groups.</li> <li>Provide the survival probabilities at specific follow-up time points (outcome free probabilities where the outcome is not death), median survival time and 95% confidence interval is preferred for comparison with other studies. This should be provided for each group as well when the objective is to compare groups.</li> <li>Report the method used to estimate the survival probabilities and plot the survival curves using appropriate graphs like Kaplan-Meier or Nelson-aalen cumulative curve stratified by groups when necessary. Include the following information in the survival curves: number of participants at risk at each specified timepoint, indicate when participants were censored, use different colors/type of lines to distinguish group curves and clearly label the x-axis as time under observation and y-axis appropriately.</li> <li>When testing hypothesis of differences in survival probabilities between/among groups, report the method used, the test results and a P-value.</li> <li>When survival regression is performed, report the methods used to test underlying assumptions (test for Proportion Hazard assumption for Cox regression and test of used probability distribution for parametric methods) and the test result.</li> <li>Report the measure of effect (e.g Hazard ratios, sub-distribution hazard ratios, time ratios), their measure of uncertainty (e.g 95% confidence intervals, standard errors) and P-values from the regression model.</li> <li>Like other statistical regression modeling, report all the covariates assessed, method of selecting features to be included in the multivariable survival regression model, methods used to assess the multivariable regression goodness of fit, proportion of missing data in the outcome and covariates assessed plus how missing data were handled, methods used to test for interaction and methods used to control for clustering in multilevel studies.</li> </ul>

this review, its likely they analyzed all the available records, but in such settings authors should be encouraged to perform a priori sample size estimation [84].

More than three quarters of the studies with some form of clustering of participants did not consider the design aspect in the analysis. This may point to a major challenge in the analysis of such designs despite there being comprehensive statistical methods of investigating cluster heterogeneity and controlling for the extra level of variation [31, 85, 86]. Not accounting for clustering in the analysis, may yield biased and extreme results leading to a false conclusion [13]. However, it was encouraging to observe, all the statistical software reported have robust systems to handle survival analyses, investigate and perform adjustments for non-informative censoring and clustering. Reporting of sample size estimation and accounting for clustering in analysis are not SAMs specific issues but the low frequency of adequacy of their reporting in this review, raises the possibility of suboptimal practices across reporting of TB in general.

The Consolidated Standards of Reporting Trials (CONSORT) and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines were developed to harmonize and improve quality of reporting randomized control trials (RCTs) and observational studies respectively, however, their focus is not on specific statistical methods [87, 88]. Recommendations on how to report specific statistical topics like missing data imputation [89], Bayesian analysis [90], and logistic regression [91] have been developed. Apart from suggestions by two previous reviews of SAMs [16, 17], currently there is no recommended standard guidelines for reporting SAMs. From our findings, we propose some pragmatic recommendations (Table 6) for researchers, statisticians and journal editors and emphasize the need to develop harmonized guideline for reporting SAMs.

Excluding non-English papers was one of the study limitations. However, looking at the countries where the studies were conducted, suggests the Francophone and other non-English speaking countries (like Ethiopia and Mozambique) were not excluded but could be underrepresented. The reporting of SAMs may be influenced by many factors like involvement of statistician/epidemiologist, but it was challenging ascertaining involvement and level of skills of the statistician/epidemiologist and the likely lack of power to perform such analysis. We thus explored the effect of such factors in sub-analysis.

## Conclusion

The quality of reported survival analyses in studies of TB in Africa is inadequate despite the increasing number of annual publications on the topic. Our findings suggest sample size estimation, testing of underlying survival

regression models and visual display of the survival function were rarely adequately reported. Some of these deficiencies may lead to incorrect results and conclusion. Because similar reporting deficiencies may be common in other diseases in low- and middle-income countries, reporting guidelines, additional training and more capacity building are needed along with more vigilance by reviewers and journal editors.

## Abbreviations

AIC: Akaike Information Criteria; ALL: Acute lymphoblastic leukemia; AUCs: Area under receiver operating characteristic curve; BIC: Bayesian Information Criteria; CPH: Cox Proportional Hazard; HIV: human immunodeficiency viruses; HR: Hazard ratios; IJTL: International Journal of Tuberculosis and Lung Diseases; KM: Kaplan-Meier; LL: Likelihood; PH: Proportional Hazard; REDCap: Research electronic data capture; RMST: Restricted mean survival time; SAM: Survival analyses methods; SHR: Subhazard ratios; TB: Tuberculosis; TR: Time ratios; WHO: World Health organization

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-021-01280-3>.

### Additional file 1.

## Acknowledgements

The authors wish to acknowledge Coralie Dessenne, the documentation officer, coordination and support at Department of Population Health, Luxembourg Institute of Health for her help with searching for the papers.

## Authors' contributions

MMN conceived the study. MMN, SS, CM, LM and MV designed the study. SS and MMN were involved in search of the papers. CM and MMN performed screening of studies and data extraction. MMN performed data analysis and writing of the first manuscript draft. MV provided overall supervision. All authors read and approved the final manuscript.

## Funding

MMN is currently supported by the WHO/TDR Clinical Research and Development Fellowships Program. The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Availability of data and materials

The datasets used and analyzed in this current study are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare they have no competing interests.

## Author details

<sup>1</sup>KEMRI/Wellcome Trust Research Programme, P.O Box 230, Kilifi 80108, Kenya. <sup>2</sup>The Childhood Acute Illness & Nutrition Network (CHAIN), Nairobi, Kenya. <sup>3</sup>Competence Center for Methodology and Statistics, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg. <sup>4</sup>Department of Biostatistics and Data Science, University of Kansas Medical Center, Kansas, USA.

Received: 9 November 2020 Accepted: 12 April 2021  
Published online: 27 April 2021

## References

- Chai-Adisaksopha C, Iorio A, Hillis C, Lim W, Crowther M. A systematic review of using and reporting survival analyses in acute lymphoblastic leukemia literature. *BMC Hematol*. 2016;16:17.
- Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer*. 2003; 89(2):232–8.
- Hosmer DW, Lemeshow S, May S. Descriptive methods for survival data. In: *Applied Survival Analysis*. 2nd ed. Hoboken: Wiley; 2008. p. 16–66.
- Bruce NG, Pope D, Stanistreet DL. Life tables, survival analysis, and Cox regression. In: *Quantitative Methods for Health Research*; . 2017. <https://doi.org/10.1002/9781118665374.ch8>
- Abd ElHafeez S, Torino C, D'Arrigo G, Bolignano D, Provenzano F, Mattace-Raso F, et al. An overview on standard statistical methods for assessing exposure-outcome link in survival analysis (part II): the Kaplan-Meier analysis and the Cox regression method. *Aging Clin Exp Res*. 2012;24(3):203–6.
- Bradburn MJ, Clark TG, Love SB, Altman DG. Survival Analysis Part II: Multivariate data analysis- An introduction to concepts and methods. *Br J Cancer*. 2003;89(3):431–6.
- Cox DR. Regression models and life tables (with discussion). *J R Stat Soc*. 1972;34:187–220.
- Schober P, Vetter TR. Survival analysis and interpretation of time-to-event data: the tortoise and the hare. *Anesth Analg*. 2018;127(3):792–8. <https://doi.org/10.1213/ANE.0000000000003653>.
- Wei LJ. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Stat Med*. 1992;11(14-15):1871–9. <https://doi.org/10.1002/sim.4780111409>.
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496–509. <https://doi.org/10.1080/01621459.1999.10474144>.
- Amorim LDAF, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol*. 2015;44(1):324–33. <https://doi.org/10.1093/ije/dyu222>.
- Austin PC. A tutorial on multilevel survival analysis: methods, models and applications. *Int Stat Rev*. 2017;85(2):185–203. <https://doi.org/10.1111/insr.12214>.
- Galbraith S, Daniel JA, Vissel B. A study of clustered data and approaches to its analysis. *J Neurosci*. 2010;30(32):10601–8. <https://doi.org/10.1523/JNEUROSCI.0362-10.2010>.
- Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995;48(12):1503–10.
- Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983;39(2):499–503. <https://doi.org/10.2307/2531021>.
- Altman DG, De Stavola BL, Love SB, Stepniowska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72(2):511–8.
- Abraira V, Muriel A, Emparanza JI, Pijoan JI, Royuela A, Plana MN, et al. Reporting quality of survival analyses in medical journals still needs improvement. A minimal requirements proposal. *J Clin Epidemiol*. 2013; 66(12):1340–6.e5.
- Rulli E, Ghilotti F, Biagioli E, Porcu L, Marabese M, D'Incalci M, et al. Assessment of proportional hazard assumption in aggregate data: a systematic review on statistical methodology in clinical trials using time-to-event endpoint. *Br J Cancer*. 2018;119(12):1456–63. <https://doi.org/10.1038/s41416-018-0302-8>.
- Batson S, Greenall G, Hudson P. Review of the reporting of survival analyses within randomised controlled trials and the implications for meta-analysis. *PLoS One*. 2016;11(5):e0154870.
- Zhu X, Zhou X, Zhang Y, Sun X, Liu H, Zhang Y. Reporting and methodological quality of survival analysis in articles published in Chinese oncology journals. *Med (United States)*. 2017;96(50):e9204.
- Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med*. 2002; 136(2):122–6. <https://doi.org/10.7326/0003-4819-136-2-200201150-00009>.
- Ndouna Diakou LA, Ntoumi F, Ravaud P, Boutron I. Avoidable waste related to inadequate methods and incomplete reporting of interventions: a systematic review of randomized trials performed in sub-Saharan Africa. *Trials*. 2017;18(1):291. <https://doi.org/10.1186/s13063-017-2034-0>.
- WHO. Global tuberculosis report 2018. Geneva: World Health Organization; 2018. p. 2018.
- Kyu HH, Maddison ER, Henry NJ, Mumford JE, Barber R, Shields C, et al. The global burden of tuberculosis: results from the global burden of disease study 2015. *Lancet Infect Dis*. 2018;18(3):261–84. [https://doi.org/10.1016/S1473-3099\(17\)30703-X](https://doi.org/10.1016/S1473-3099(17)30703-X).
- WHO. Definitions and reporting framework for tuberculosis – 2013 revision. 2014.
- Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Rev Esp Nutr Humana y Diet*. 2016;20:148–60.
- London S, Gurdal O, Gall C. Automatic export of PubMed® citations to EndNote®. *Med Ref Serv Q*. 2010;29(2):146–53. <https://doi.org/10.1080/02763861003723317>.
- Elmagarmid A, Fedorowicz Z, Hammady H, Ilyas I, Khabisa M, Ouzzani M. Rayyan: a systematic reviews web app for exploring and filtering searches for eligible studies for Cochrane Reviews. In: *Evidence-Informed Public Health: Opportunities and Challenges*. Abstracts of the 22nd Cochrane Colloquium; 2014.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–81. <https://doi.org/10.1016/j.jbi.2008.08.010>.
- Bradburn MJ, Clark TG, Love SB, Altman DG. Survival Analysis Part III: Multivariate data analysis - Choosing a model and assessing its adequacy and fit. *Br J Cancer*. 2003;89(4):605–11.
- Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part IV: further concepts and methods in survival analysis. *Br J Cancer*. 2003;89(5):781–6.
- Cuzick J. A wilcoxon-type test for trend. *Stat Med*. 1985;4(4):543–7. <https://doi.org/10.1002/sim.4780040416>.
- Brust JCM, Shah NS, Mlisana K, Moodley P, Allana S, Campbell A, et al. Improved Survival and Cure Rates with Concurrent Treatment for Multidrug-Resistant Tuberculosis-Human Immunodeficiency Virus Coinfection in South Africa. *Clin Infect Dis*. 2018;66(8):1246–53.
- Azeez A, Ndege J, Mutambayi R. Associated factors with unsuccessful tuberculosis treatment outcomes among tuberculosis/HIV coinfecting patients with drug-resistant tuberculosis. *Int J Mycobacteriol*. 2018;7(4):347–54. [https://doi.org/10.4103/ijmy.ijmy\\_140\\_18](https://doi.org/10.4103/ijmy.ijmy_140_18).
- Dhedda K, Shean K, Zumla A, Badri M, Streicher EM, Page-Shipp L, et al. Early treatment outcomes and HIV status of patients with extensively drug-resistant tuberculosis in South Africa: a retrospective cohort study. *Lancet*. 2010;375(9728):1798–807. [https://doi.org/10.1016/S0140-6736\(10\)60492-8](https://doi.org/10.1016/S0140-6736(10)60492-8).
- Acuña-Villaorduña C, Ayakaka I, Dryden-Peterson S, Nakubulwa S, Worodria W, Reilly N, et al. High mortality associated with retreatment of tuberculosis in a clinic in Kampala, Uganda: A retrospective study. *Am J Trop Med Hyg*. 2019;101(1):73–5.
- Gupta-Wright A, Fielding K, Wilson D, van Oosterhout JJ, Grint D, Mwandumba HC, et al. Tuberculosis in hospitalized patients with human immunodeficiency virus: clinical characteristics, mortality, and implications from the rapid urine-based screening for tuberculosis to reduce AIDS related mortality in hospitalized patients in Africa. *Clin Infect Dis*. 2020; 71(10):2618–26.
- Zetola NM, Modongo C, Moonan PK, Ncube R, Matlhagela K, Sepako E, et al. Clinical outcomes among persons with pulmonary tuberculosis caused by *Mycobacterium tuberculosis* isolates with phenotypic heterogeneity in results of drug-susceptibility tests. *J Infect Dis*. 2014;209(11):1754–63.
- Daniels JF, Khogali M, Mohr E, Cox V, Moyo S, Edginton M, et al. Time to ART initiation among patients treated for rifampicin-resistant tuberculosis in khayelitsha, South Africa: Impact on mortality and treatment success. *PLoS One*. 2015;10(11):e0142873.
- Gesesew H, Tsehayneh B, Massa D, Gebremedhin A, Kahsay H, Mwanri L. Predictors of mortality in a cohort of tuberculosis/HIV co-infected patients in Southwest Ethiopia. *Infect Dis Poverty*. 2016;5(1):109.
- Marx FM, Dunbar R, Enarson DA, Beyers N. The rate of sputum smear-positive tuberculosis after treatment default in a high-burden setting: a retrospective cohort study. *PLoS One*. 2012;7(9):e45724. <https://doi.org/10.1371/journal.pone.0045724>.
- Getachew T, Bayray A, Weldearegay B. Survival and predictors of mortality among patients under multi-drug resistant tuberculosis treatment in Ethiopia: St. Peter's specialized tuberculosis hospital, Ethiopia. *Int J Pharm*

- Sci Res. 2013;4(2):776–87. [https://doi.org/10.13040/IJPSR.0975-8232.4\(2\).776-87](https://doi.org/10.13040/IJPSR.0975-8232.4(2).776-87).
43. Pepper DJ, Schomaker M, Wilkinson RJ, Azevedo V, Maartens G. Independent predictors of tuberculosis mortality in a high HIV prevalence setting: a retrospective cohort study. *AIDS Res Ther.* 2015;12(1):35. <https://doi.org/10.1186/s12981-015-0076-5>.
  44. Hafkin J, Modongo C, Newcomb C, Lowenthal E, MacGregor RR, Steenhoff AP, et al. Impact of the human immunodeficiency virus on early multidrug-resistant tuberculosis treatment outcomes in Botswana. *Int J Tuberc Lung Dis.* 2013;17(3):348–53.
  45. Abdullahi OA, Ngari MM, Sanga D, Katana G, Willetts A. Mortality during treatment for tuberculosis; a review of surveillance data in a rural county in Kenya. *PLoS One.* 2019;14(7):e0219191. <https://doi.org/10.1371/journal.pone.0219191>.
  46. Hueriga H, Ferlazzo G, Wanjala S, Bastard M, Bevilacqua P, Ardizzone E, et al. Mortality in the first six months among HIV-positive and HIV-negative patients empirically treated for tuberculosis. *BMC Infect Dis.* 2019;19(1):132.
  47. Kassa GM, Teferra AS, Wolde HF, Muluneh AG, Merid MW. Incidence and predictors of lost to follow-up among drug-resistant tuberculosis patients at University of Gondar Comprehensive Specialized Hospital, Northwest Ethiopia: A retrospective follow-up study. *BMC Infect Dis.* 2019;19(1):817.
  48. Yihunie Akalu T, Fentahun Muchie K, Alemu GK. Time to sputum culture conversion and its determinants among multi-drug resistant tuberculosis patients at public hospitals of the Amhara regional state: a multicenter retrospective follow up study. *PLoS One.* 2018;13(6):e0199320. <https://doi.org/10.1371/journal.pone.0199320>.
  49. Ketema DB, Muchie KF, Andargie AA. Time to poor treatment outcome and its predictors among drug-resistant tuberculosis patients on second-line anti-tuberculosis treatment in Amhara region, Ethiopia: Retrospective cohort study. *BMC Public Health.* 2019;19(1):1481.
  50. Limenih YA, Workie DL. Survival analysis of time to cure on multi-drug resistance tuberculosis patients in Amhara region, Ethiopia. *BMC Public Health.* 2019;19(1):165.
  51. Wickett E, Peralta-Santos A, Beste J, Micikas M, Toe F, Rogers J, et al. Treatment outcomes of TB-infected individuals attending public sector primary care clinics in rural Liberia from 2015 to 2017: a retrospective cohort study. *Trop Med Int Heal.* 2018;23(5):549–57. <https://doi.org/10.1111/tmi.13049>.
  52. Farley JE, Ram M, Pan W, Waldman S, Cassell GH, Chaisson RE, et al. Outcomes of multi-drug resistant tuberculosis (MDR-TB) among a cohort of south African patients with high HIV prevalence. *PLoS One.* 2011;6(7):e20436. <https://doi.org/10.1371/journal.pone.0020436>.
  53. Alene KA, Viney K, McBryde ES, Tsegaye AT, Clements ACA. Treatment outcomes in patients with multidrug-resistant tuberculosis in north-west Ethiopia. *Trop Med Int Heal.* 2017;22(3):351–62.
  54. Azeez A, Mutambayi R, Odeyemi A, Ndege J. Survival model analysis of tuberculosis treatment among patients with human immunodeficiency virus coinfection. *Int J Mycobacteriol.* 2019;8(3):244–51. [https://doi.org/10.4103/ijmy.ijmy\\_101\\_19](https://doi.org/10.4103/ijmy.ijmy_101_19).
  55. Onyango DO, Yuen CM, Masini E, Borgdorff MW. Epidemiology of pediatric tuberculosis in Kenya and risk factors for mortality during treatment: a National Retrospective Cohort Study. *J Pediatr.* 2018;201:115–21. <https://doi.org/10.1016/j.jpeds.2018.05.017>.
  56. Schutz C, Barr D, Andrade BB, Shey M, Ward A, Janssen S, et al. Clinical, microbiologic, and immunologic determinants of mortality in hospitalized patients with HIV-associated tuberculosis: a prospective cohort study. *PLoS Med.* 2019;16(7):e1002840. <https://doi.org/10.1371/journal.pmed.1002840>.
  57. Schwoebel V, Trébucq A, Kashongwe Z, Bakayoko AS, Kuaban C, Noeske J, et al. Outcomes of a nine-month regimen of rifampicin-resistant tuberculosis up to 24 months after treatment completion in nine African countries. *EClinicalMedicine.* 2020;20:100268. <https://doi.org/10.1016/j.eclinm.2020.100268>.
  58. Worodria W, Ssempijja V, Hanrahan C, Ssegonja R, Muhofwa A, Mazapkwé D, et al. Opportunistic diseases diminish the clinical benefit of immediate co-infected antiretroviral adults with therapy low CD4 in HIV-tuberculosis R cell counts. *AIDS.* 2018;32(15):2141–9. <https://doi.org/10.1097/QAD.0000000000001941>.
  59. Henegar CE, Behets F, Vanden Driessche K, Tabala M, Bahati E, Bola V, et al. Mortality among tuberculosis patients in the Democratic Republic of Congo. *Int J Tuberc Lung Dis.* 2012;16(9):1199–204. <https://doi.org/10.5588/ijtld.11.0613>.
  60. Mupere E, Malone L, Zalwango S, Chiunda A, Okwera A, Parraga I, et al. Lean Tissue Mass Wasting is Associated With Increased Risk of Mortality Among Women With Pulmonary Tuberculosis in Urban Uganda. *Ann Epidemiol.* 2012;22(7):466–73.
  61. Kendall EA, Theron D, Franke MF, Van Helden P, Victor TC, Murray MB, et al. Alcohol, hospital discharge, and socioeconomic risk factors for default from multidrug resistant tuberculosis treatment in rural South Africa: a retrospective cohort study. *PLoS One.* 2013;8(12):e83480. <https://doi.org/10.1371/journal.pone.0083480>.
  62. Onyango DO, Yuen CM, Cain KP, Ngari F, Masini EO, Borgdorff MW. Reduction of HIV-associated excess mortality by antiretroviral treatment among tuberculosis patients in Kenya. *PLoS One.* 2017;12(11):e0188235.
  63. Poorolajal J, Cheraghi Z, Irani AD, Rezaeian S. Quality of cohort studies reporting post the strengthening the reporting of observational studies in epidemiology (STROBE) statement. *Epidemiol Health.* 2011;33. <https://doi.org/10.4178/epih/e2011005>.
  64. Aghazadeh-Attari J, Mobaraki K, Ahmadzadeh J, Mansorian B, Mohebbi I. Quality of observational studies in prestigious journals of occupational medicine and health based on strengthening the reporting of observational studies in epidemiology (STROBE) statement: a cross-sectional study. *BMC Res Notes.* 2018;11(1):266. <https://doi.org/10.1186/s13104-018-3367-9>.
  65. Kim KH, Kang JW, Lee MS, Lee JD. Assessment of the quality of reporting in randomised controlled trials of acupuncture in the Korean literature using the CONSORT statement and STRICTA guidelines. *BMJ Open.* 2014;4(7):e005068.
  66. Janackovic K, Puljak L. Reporting quality of randomized controlled trial abstracts in the seven highest-ranking anesthesiology journals. *Trials.* 2018;19(1):591.
  67. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet.* 2002;359(9318):1686–9. [https://doi.org/10.1016/S0140-6736\(02\)08594-X](https://doi.org/10.1016/S0140-6736(02)08594-X).
  68. Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika.* 1981;68(1):316–9. <https://doi.org/10.1093/biomet/68.1.316>.
  69. Yang S, Prentice R. Improved logrank-type tests for survival data using adaptive weights. *Biometrics.* 2010;66(1):30–8.
  70. Mantel N. Chi-Square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *J Am Stat Assoc.* 1963;58(303):690–700. <https://doi.org/10.1080/01621459.1963.10500879>.
  71. Zucker DM, Lakatos E. Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika.* 1990;77(4):853–864.
  72. Breslow N. A generalized Kruskal-Wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika.* 1970;57(3):579–94. <https://doi.org/10.1093/biomet/57.3.579>.
  73. Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. *Biometrika.* 1982;69(3):553–66. <https://doi.org/10.2307/2335991>.
  74. Tarone RE, Ware J. On distribution-free tests for equality of survival distributions. *Biometrika.* 1977;64(1):156–60. <https://doi.org/10.1093/biomet/64.1.156>.
  75. Nardi A, Schemper M. Comparing Cox and parametric models in clinical studies. *Stat Med.* 2003;22(23):3597–610.
  76. Zare A, Hosseini M, Mahmoodi M, Mohammad K, Zeraati H, Holakouie NK. A comparison between accelerated failure-time and cox proportional hazard models in analyzing the survival of gastric cancer patients. *Iran J Public Health.* 2015;44(8):1095–102.
  77. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med.* 2011;30(19):2409–21. <https://doi.org/10.1002/sim.4274>.
  78. Templeton AJ, Amir E, Tannock IF. Informative censoring — a neglected cause of bias in oncology trials. *Nat Rev Clin Oncol.* 2020;17(6):327–8. <https://doi.org/10.1038/s41571-020-0368-0>.
  79. Schuster NA, Hoogendijk EO, Kok AAL, Twisk JWR, Heymans MW. Ignoring competing events in the analysis of survival data may lead to biased results: a nonmathematical illustration of competing risk analysis. *J Clin Epidemiol.* 2020;122:42–8. <https://doi.org/10.1016/j.jclinepi.2020.03.004>.
  80. Gray RJ. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *Ann Stat.* 1988;16:1141–54. <https://www.jstor.org/stable/2241622>.

81. Dey T, Mukherjee A, Chakraborty S. A Practical Overview and Reporting Strategies for Statistical Analysis of Survival Studies. *Chest*. 2020;158(15):S39–48.
82. Brooks MB, Mitnick CD, Manjourides J. Comparison of censoring assumptions to reduce bias in tuberculosis treatment cohort analyses. *PLoS One*. 2020;15(10):e0240297.
83. Brooks MB, Keshavjee S, Gelmanova I, Zemlyanaya NA, Mitnick CD, Manjourides J. Use of predicted vital status to improve survival analysis of multidrug-resistant tuberculosis cohorts. *BMC Med Res Methodol*. 2018; 18(1):166.
84. Johnston KM, Lakzadeh P, BMK D, Szabo SM. Methods of sample size calculation in descriptive retrospective burden of illness studies. *BMC Med Res Methodol*. 2019;19(1):9.
85. Balakrishnan N, Peng Y. Generalized gamma frailty model. *Stat Med*. 2006; 25(16):2797–816.
86. O'Quigley J, Stare J. Proportional hazards models with frailties and random effects. *Stat Med*. 2002;21(21):3219–33.
87. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340(mar23 1):c869. <https://doi.org/10.1136/bmj.c869>.
88. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12(12):1495–9.
89. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ (Online)*. 2009;338:b2393.
90. Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. *J Clin Epidemiol*. 2005;58(3):261–8.
91. Kalil AC, Mattei J, Florescu DF, Sun J, Kalil RS. Recommendations for the assessment and reporting of multivariable logistic regression in transplantation literature. *Am J Transplant*. 2010;10(7):1686–94.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

