





MATHEMATICS, STATISTICS AND PROBABILITY
SUPPLEMENTARY RESULT

Addressing the statistical analysis dilemma that exists when analyzing clinical trial results with full efficacy using the Kaplan Meier survival analysis method

Pimnara Peerawaranun¹ , Rob W. van der Pluijm^{1,2} and Mavuto Mukaka^{1,2,*} 

¹Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, and

²Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, United Kingdom

*Corresponding author. Email: mmukaka@gmail.com

(Received 07 June 2021; Revised 16 September 2021; Accepted 17 September 2021)

Abstract

The use of a Kaplan–Meier (K–M) survival time approach is generally considered appropriate to report antimalarial efficacy trials. However, when a treatment arm has 100% efficacy, confidence intervals may not be computed. Furthermore, methods that use probability rules to handle missing data for instance by multiple imputation, encounter perfect prediction problem when a treatment arm has full efficacy, in which case all imputed values are either treatment success or all imputed values are failures. The use of a survival K–M method addresses this imputation problem in estimating the efficacy estimates also referred to as cure rates. We discuss the statistical challenges and propose a potential way forward.

The proposed approach includes the use of K–M estimates as the main measure of efficacy. Confidence intervals could be computed using the binomial exact method. *p*-Values for comparison of difference in efficacy between treatments can be estimated using Fisher's exact test. We emphasize that when efficacy rates are not 100% in both groups, the K–M approach remains the main strategy of analysis considering its statistical robustness in handling missing data and confidence intervals can be computed under such scenarios.

Key words: efficacy; intention to treat; Kaplan–Meier; missing data; multiple imputation

1. Background

Randomized controlled clinical trials are often used to assess treatment efficacy or to compare efficacy between treatment groups (Montori & Guyatt, 2001). The procedure involves randomizing participants to the different treatments and then follow up the participants for a set period of time. At the end of the study period, the efficacy of the treatment is determined. Efficacy is defined as the proportion (or percentage) of subjects that have a treatment success out of the total population exposed to the treatment arm.

There are many ways of analyzing and reporting efficacy data. The main outcome measures include the risk differences, which are simply the differences in the efficacy proportions; the odds ratios; the risk ratios and the hazard ratios (Magder, 2003; Mukaka et al., 2016). In order to allow for comparisons with previous studies, it is important that the measures of effect are consistent with existing literature and/or comply with the requirements of regulatory authorities. It is common practice to use survival models to report cure rates, especially in antimalarial efficacy trials (Dahal et al., 2019). The main advantage of the

survival methods is that they can handle missing data without completely discarding the subject that for which data missing. Thus, this approach increases the precision of the estimates (Dahal et al., 2017). In this way subjects that are lost to follow up; withdraw consent; represent with and retreated for reinfection and/or different malaria species, still contribute to the estimation of treatment efficacy up to the moment of being censored in the Kaplan–Meier (K–M) survival analysis. The analysis of data involves both the univariate analyses and the multivariable analyses. The Cox Proportional Hazards model is often used for the multivariable analysis to complement the K–M method which is mainly used for the univariate efficacy estimates. The univariate estimates of efficacy are often justified in randomized trials considering that randomization balances both known and unknown confounders among the treatment groups.

The efficacy estimates can easily be obtained as proportions along with the confidence intervals using the binomial exact calculation method. This method works for both 0% efficacy and 100% efficacy. The main shortfall of the proportion approach is that it fails to handle censored outcomes, that is, observation with partial information about the outcome. For this reason the use of the K–M method becomes appealing.

2. Statistical analysis dilemma

When efficacy of a treatment is 100% or 0%, there are statistical challenges with other approaches such as the survival K–M method. For example, the confidence intervals cannot be computed if the survival K–M methods are used to estimate cure rates from the treatment with 100% efficacy as is the case with p -values. The absence of the confidence intervals and p -values make it difficult to perform a hypothesis test comparing two or more treatment groups. For this reason, we need an alternative method for computing the confidence intervals and the corresponding p -values while still reporting the K–M estimates of efficacy which help in preserving sample size through censoring of individuals with incomplete follow-up information.

Let $S(t)$ be the survival function that can be used to estimate the cure rates, then the 95% confidence interval (that is commonly used) is given by (Collett, 2004)

$$\widehat{S}(t) \exp \left[\pm \frac{1.96}{\ln S(t)} \frac{s \cdot e}{S(t)} \right], \quad (1)$$

where $s \cdot e \approx \widehat{S}(t) \sqrt{\sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}}$ is the standard error of the survival probability, $S(t)$, at any time t , $t_k \leq t < t_{k+1}$, n_j is the number of participants who are at risk at time j , d_j are the number of events (treatment failures) observed between time $j - 1$ and j .

When there are no failures, then survival probability $S(t)$ is 1 and $\ln S(t)$ becomes 0. With full efficacy, substituting $\ln S(t) = \ln(1) = 0$ in the expression (1) above, $\frac{1.96}{\ln S(t)}$, results in undefined confidence interval as $\frac{1.96}{0}$ is undefined. Furthermore, when there are no failures, the standard error of the survival probability, $S(t)$ which is given by $S(t) \sqrt{\sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}}$ reduces to zero. Hence, the confidence interval for the cure rate estimated by the K–M may not be computed when either there is full efficacy or zero efficacy in the study sample.

In order to illustrate the advantage of using the survival K–M method, we consider its application in the Intention to treat (ITT) analyses. In the ITT analysis, all subjects randomized to a specific treatment arm are included into the analysis, irrespective of whether or not the subjects encounter any of competing events such as reinfection, were lost to follow up or withdraw consent. In the per protocol analysis subjects that presented with a reinfection, were lost to follow up or withdraw consent, are excluded. In malaria trials specifically, the incidence of ‘competing events’ can be considerable. For instance, in high transmission areas reinfection rates could be as high as 25%.

For the ITT analysis, a potential approach for dealing with missing data is to perform multiple imputation (MI), use the inverse probability weighting (IPW) or the doubly robust-IPW (DR-IPW) method (Machekano et al., 2008). In this approach the outcome of the subjects that encountered ‘competing events’ is predicted based on the outcome of the subjects that reached the endpoint (either completion of follow up or treatment failure). However, if treatment efficacy is 100% or 0% there is a perfect prediction problem for the outcome, and all subjects with missing outcomes will be imputed as treatment success or treatment failure, respectively. This is where the use of survival K–M method to estimate the efficacy is an important tool. It immediately addresses the need to worry with imputing outcomes for those participants that have missing data, hence helps in preserving the sample sizes for the ITT principle.

The K–M Survival estimates, MI, IPW and DR-IPW strategies work optimally in scenarios where efficacy is not 100% or 0%. When analyzing data with 100% or 0% efficacy, the analysis and reporting of such data results in a significant statistical dilemma when missing outcome data exists. We focus on how to address this problem when applying the K–M method to estimate the cure rates.

3. Here is a hypothetical example: Data simulation and analysis results

The data was simulated in excel (Supplementary appendix) and imported into Stata 16 for analysis. Hypothetically, 200 participants are randomized into two arms of a Malaria efficacy trial. The primary endpoint is defined as the day 42 PCR corrected efficacy. The data set for demonstration was simulated as follows. A set of 100 individuals containing four variables ID, treatment (0 or 1), outcome (0 = success, 1 = failure) and time which was day of outcome measurement were simulated for one arm. There are no failures in this treatment arm with 100 individuals by the end of the trial follow-up period. The outcome variable was recorded as a 0 for all individuals in this arm. Another set of 100 individuals was simulated as the second treatment arm. In this arm, there are five failures with three individuals failing at days 14, 21, 35 respectively and two individuals failing at day 28 (Supplementary appendix). As shown in Table 1, treatment 0, the K–M estimate of the day 42 cure rate (with 95%CI) is 95.0% (88.4%, 97.9%). On the other hand, in treatment 1, the K–M estimate of the day 42 cure rate is not estimated because the cure rate is 100%.

Table 1. Day 42 efficacy estimates by treatment group using the Kaplan–Meier survival method

.sts list, by (treatment)							
failure _d: outcome							
analysis time _t: time							
Time	Beg. total	Fail	Net lost	Survivor function	Standard error	[95% confidence interval]	
Treatment = 0							
14	100	1	0	0.99	0.01	0.93	1.00
21	99	1	0	0.98	0.01	0.92	1.00
28	98	2	0	0.96	0.02	0.90	0.98
35	96	1	0	0.95	0.02	0.88	0.98
42	95	0	95	0.95	0.02	0.88	0.98
Treatment = 1							
42	100	0	100	1	–	–	–

Note: The analysis was done in Stata 16.

Table 2. Day 42 efficacy estimates by treatment group using the binomial exact calculation method

.cii proportions 100,100					
Variable	Obs.	Proportion	Standard error	Binomial exact	
				[95% confidence interval]	
	100	1	0	0.9637833	1 ^a

Note: The analysis is done in Stata 16.

^aOne-sided, 97.5% confidence interval.

4. Proposed analysis approaches

The following may be plausible ways of addressing this analysis dilemma: (a) In order to address the problem faced when using the K–M for samples with 100% cure rate, we could report the K–M efficacy estimates as the main measure to be reported. Since the K–M estimates will be 100%, it remains true that efficacy estimate from the K–M will be exactly the same as that of the simple proportion. Therefore, the binomial exact confidence intervals provide valid confidence intervals for the K–M cure rates. Thus, the K–M efficacy estimate could be reported together with the binomial exact confidence interval. The advantage of this is that the binomial exact methods can be used to obtain the confidence intervals for the efficacy when the cure rate is 100% or 0%.

Now if we go back to the hypothetical study above and use the binomial exact method to get the day 42 cure rates, the efficacy is 100% 95% (96.4%, 100%) for treatment 1, (Table 2).

Our recommendation would be to use this approach as much as possible. (b) The second approach may be to use the K–M method to estimate efficacy without associated confidence intervals as the confidence intervals cannot be computed. Comparisons based on *p*-values may be performed using the different variations of the Logrank test. The unfortunate thing about this second approach is that uncertainty of the efficacy estimates is not reflected due to missing confidence intervals. In addition, the Logrank test is not robust in this case for the reasons given above compared to using Fisher's exact test.

5. Discussion

Analysis of binary outcome data to compare the outcome between treatment arms is often problematic when one or both of the arms have either 100% or 0% success rate. When the outcome of interest is time to event, there are mathematical and computational challenges associated with computation of both the *p*-values and the confidence intervals for the difference in hazards. Hence, reporting of the results from such a study becomes complicated as the comparisons would be made descriptively without a supporting statistics.

A potential quick solution is to alter the outcome of one or more of the subjects to either failure or success in case of 100% and 0% efficacy, respectively. However, such a solution does not recognize the fact that sample treatment efficacy was in fact 100% or 0% in the participants that were assessed. Thus, such a solution prohibits researchers from making correct inferences. We suggest that when the cure rate is 100% (or 0%), the K–M estimates could still be calculated and presented as the K–M. This helps in retaining the original sample size by censoring individuals that are only partially observed. Once the cure rates have been estimated using the K–M method, the confidence could be calculated using the binomial exact method. Authors should be transparent by clearly stating this in the statistical analysis plan as well as providing relevant footnotes in the manuscripts on that has been reported and the reasons for reporting those measures.

This approach will be helpful because the inferences will be backed up by statistical evidence. Routine has been just to describe the results when there is this type of dilemma. Thus, article will therefore provide guidance when the cure rate is 100% (or 0%) and data is analyzed as time to event.

6. Conclusion

Use of the K–M method to estimate efficacy lands into analysis problems when treatment has full efficacy. When the cure rate is 100%, the K–M estimates could still be calculated and presented as the K–M estimates preserves the sample size by censoring individuals that are only partially observed. The confidence could be calculated using the binomial exact method.

Funding. This review was funded in whole, or in part, by the Wellcome Trust [220211]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Conflict of Interest. The authors declare no competing interests.

Data Availability Statement. We did not use patient data for this review article. The simulated data that we have used for the hypothetical examples has been provided in the supplementary appendix.

Supplementary Materials. To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/exp.2021.21>.

References

- Collett, D. (2004). *Modelling survival data in medical research*. Chapman and Hall; CRC.
- Dahal, P., Guerin, P. J., Price, R. N., Simpson, J. A., & Stepniewska, K. (2019). Evaluating antimalarial efficacy in single-armed and comparative drug trials using competing risk survival analysis: A simulation study. *BMC Medical Research Methodology*, *19*, 107.
- Dahal, P., Simpson, J. A., Dorsey, G., Guérin, P. J., Price, R. N., & Stepniewska, K. (2017). Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated *Plasmodium falciparum* malaria: Pitfalls and challenges. *Malaria Journal*, *16*, 430.
- Machekano, R. N., Dorsey, G., & Hubbard, A. (2008). Efficacy studies of malaria treatments in Africa: Efficient estimation with missing indicators of failure. *Statistical Methods in Medical Research*, *17*, 191–206.
- Magder, L. S. (2003). Simple approaches to assess the possible impact of missing outcome information on estimates of risk ratios, odds ratios, and risk differences. *Controlled Clinical Trials*, *24*, 411–421.
- Montori, V. M., & Guyatt, G. H. (2001). Intention-to-treat principle. *Canadian Medical Association Journal*, *165*, 1339–1341.
- Mukaka, M., White, S. A., Terlouw, D. J., Mwapasa, V., Kalilani-Phiri, L., & Faragher, E. B. (2016). Is using multiple imputation better than complete case analysis for estimating a prevalence (risk) difference in randomized controlled trials when binary outcome observations are missing?. *Trials*, *17*, 341.

Cite this article: Peerawaranun P, van der Pluijm RW, Mukaka M (2021). Addressing the statistical analysis dilemma that exists when analyzing clinical trial results with full efficacy using the Kaplan Meier survival analysis method *Experimental Results*, *2*, e32, 1–9. <https://doi.org/10.1017/exp.2021.21>

Peer Reviews


Reviewing editor: Dr. Brian Williamson

University of Bolton, Mathematics, School of Engineering, University of Bolton, Bolton, United Kingdom of Great Britain and Northern Ireland, BL3 5AB

This article has been accepted because it is deemed to be scientifically sound, has the correct controls, has appropriate methodology and is statistically valid, and has been sent for additional statistical evaluation and met required revisions.

doi:10.1017/exp.2021.21.pr1

Review 1: Addressing the statistical analysis dilemma that exists when analyzing clinical trial results with full efficacy using the Kaplan Meier survival analysis method.

Reviewer: Dr. Mohammad Reza Mahmoudi 

Date of review: 03 August 2021

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Conflict of interest statement. Reviewer declares none.

Comments to the Author: This is an interesting contribution to the existing literature, but the paper suffers from several shortcomings listed in the following comments.

-The paper should be checked by a native.

-A discussion section should be added.

-The introduction should be updated by recent researches.

-The novelty and contribution should be clearly bolded.

-The authors should consider some works about Data Analysis that can be applied to model different datasets. For example,

Comparison of the climate indices based on the relationship between yield loss of rain-fed winter wheat and changes of climate indices using GEE model, *Science of The Total Environment* 661, 711-722

On the detection and estimation of the simple harmonizable processes, *Iranian Journal of Science and Technology (Sciences)* 39 (2), 239-242

Two-piece location-scale distributions based on scale mixtures of normal family, *Communications in Statistics-Theory and Methods* 46 (24), 12356-12369

Large Sample Inference about the Ratio of Means in Two Independent Populations, *Journal of Statistical Theory and Applications* 16 (3), 366-374

On comparing and classifying several independent linear and non-linear regression models with symmetric errors, *Symmetry* 11 (6), 820

It's better to suggest some subjects for future works.

Best regards,

Score Card

Presentation



Is the article written in clear and proper English? (30%)

4/5

Is the data presented in the most useful manner? (40%)

4/5

Does the paper cite relevant and related articles appropriately? (30%)

4/5

Context



Does the title suitably represent the article? (25%)

4/5

Does the abstract correctly embody the content of the article? (25%)

4/5

Does the introduction give appropriate context? (25%)

4/5

Is the objective of the experiment clearly defined? (25%)

4/5

Analysis



Does the discussion adequately interpret the results presented? (40%)

4/5


Is the conclusion consistent with the results and discussion? (40%)

4/5

Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)

4/5

Review 2: Addressing the statistical analysis dilemma that exists when analyzing clinical trial results with full efficacy using the Kaplan Meier survival analysis method.

Reviewer: Dr. Marius Iulian Mihailescu 

Spiru Haret University, Engineering and Computer Science, Bucuresti, Romania, 030045

Date of review: 30 August 2021

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Conflict of interest statement. Reviewer declares none.

Comments to the Author: The author(s) should consider the following comments in an objective manner.

- The tests given in the paper provides some technical gaps.
- The gaps are:
 - o The data set is not well provided and is described accordingly.
 - o Which are the variables?
 - o Can a data set be verified somewhere? It is online? Can be checked by the reader?
 - o The reader is able to access for his own purpose the data set?
 - o What software did you use to get the coefficients? Did you R System, STATISTICA, IBM SPSS, Python, or other? Please describe and provide the scripts or pseudocode if it is possible, also a GitHub repository is welcomed as well.
- The bibliography is outdated.
 - o The most recent is from 2016
 - o Can it be updated with some new references?

Score Card

Presentation



Is the article written in clear and proper English? (30%)

4/5

Is the data presented in the most useful manner? (40%)

3/5

Does the paper cite relevant and related articles appropriately? (30%)

3/5

Context



Does the title suitably represent the article? (25%)

4/5

Does the abstract correctly embody the content of the article? (25%)

5/5

Does the introduction give appropriate context? (25%)

3/5

Is the objective of the experiment clearly defined? (25%)

3/5

Analysis



Does the discussion adequately interpret the results presented? (40%)

3/5

Is the conclusion consistent with the results and discussion? (40%)

3/5

Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)

3/5