SN Comprehensive Clinical Medicine (2023) 5:24
https://doi.org/10.1007/s42399-022-01367-y

REVIEW

Survival Analysis and Cox Proportional Hazards Model Reporting in Pediatric Leukemia Studies—a Systematic Review

Ilari Kuitunen[1,2] · Atte Nikkilä[3,4] · Ville T. Ponkilainen[5] · Mikko M. Uimonen[4] · Olli Lohi[4,6]

Accepted: 11 December 2022

Abstract

Survival (overall, event free, etc.) is the most-used outcome in clinical oncology studies. This study analyzed methodological reporting of survival analysis in pediatric leukemia studies, focusing on Cox proportional hazards (PH). We performed a systematic review of studies published between 2012 and 2021 in the five highest-ranking oncology and five highest-ranking hematology journals. The included studies had to focus on pediatric leukemia and utilize survival analyses. We extracted data on how the survival analysis methodology was reported and focused on Cox proportional hazards modeling and whether the PH assumption was checked. We screened 561 studies and included 103 in the analysis. The most-used crude survival analysis method was Kaplan–Meier, as 96 (94%) of the 103 studies applied it. Adjusted survival analysis was performed in 80 (78%) of the included studies, and the Cox PH model was used in 77 (96%) of these studies. The PH assumption was mentioned in 18 (23%) of the 77 studies that used the Cox PH model. Only nine studies (12%) stated how the PH assumption was assessed. We noted 10 (13%) studies with possible violations of the PH assumption. Overall, we found a need for improvement in the reporting of survival analysis and especially PH assumption in pediatric leukemia studies. The Cox PH model was the most-used adjusted survival analysis method but checking of the background assumption was not reported in most of the studies.

Keywords Survival analysis · Statistics · Regression analysis · Proportional hazards

Introduction

Survival is a key outcome measure in oncological studies, and it is defined as the time from exposure to event [1–3]. The most common outcomes are overall survival, event-free survival, disease-free survival, relapse-free survival, and progression-free survival. The most-used crude survival estimation method is the Kaplan–Meier analysis, while the adjusted time-to-event analysis most often applied is the Cox proportional hazards (PH) regression model [4, 5].

The Cox PH model was introduced in 1972 and has since been among the most used survival analysis methods [6]. The most important background assumption in the Cox PH model is the proportional hazards assumption, meaning that the covariate-related hazard stays proportional over time [7]. Therefore, violation of the PH assumption may lead to biased estimates, as the hazard at one time point may not be equal to time-varying hazard at other time points [8]. The PH assumption can be tested in multiple ways, including the use of log-minus-log plots or scaled Schoenfeld residuals [9]. If the PH assumption is not met, the Cox PH model can be modified to meet the assumption in multiple ways, for example by using time-dependent coefficients or time-axis division [7, 9].

The Cox PH model has been criticized due to the mathematical and impractical nature of the proportionality assumption, which rarely holds in real life [10]. Therefore,

This article is part of the Topical Collection on *Medicine*

✉ Ilari Kuitunen
ilari.kuitunen@uef.fi

1 Institute of Clinical Medicine and Department of Pediatrics, University of Eastern Finland, Kuopio, Finland

2 Department of Pediatrics, Mikkeli Central Hospital, Porrassalmenkatu 35-37, 50100 Mikkeli, Finland

3 Department of Pediatrics, Kanta-Häme Central Hospital, Hämeenlinna, Finland

4 Tampere Center for Child, Adolescent and Maternal Health Research, Tampere University, Tampere, Finland

5 Department of Surgery, Central Finland Central Hospital Nova, Jyväskylä, Finland

6 Department of Pediatrics, Tampere University Hospital and Tays Cancer Centre, Tampere, Finland

Published online: 19 December 2022

it has been discussed that the Cox model should be replaced, for example, by a restricted mean survival time analysis [11, 12]. Alternatively, in certain conditions, authors may choose not to test the PH assumption, but this choice should be justified in the manuscript [8, 10].

The desired goal for medical research articles is reproducibility based on reporting by the authors. To this end, reporting guidelines have been created to uniform and improve the level of reporting [13, 14]. However, there is an ongoing issue with poor reproducibility in medical science, with insufficient reporting of the methods used [15, 16].

The aim of this study was to evaluate survival analysis methods and their reporting, focusing on the Cox proportional hazards model and proportional hazards assumption in pediatric leukemia studies.

Materials

We conducted a systematic methodological review of survival analysis methods and their reporting in pediatric leukemia studies.

Search Strategy

A search strategy was created with the help of two informaticians from the Library of the University of Eastern Finland. We decided to focus on the five highest-ranked clinical oncology and five clinical hematology journals, as we expected these to have high-quality reporting and that the studies they published have had the highest influence on treatments. We ranked the journals based on their impact factors in 2020, which were obtained from Clarivate analytics. The included journals were Lancet Oncology, Journal of Clinical Oncology, Annals of Oncology, JAMA Oncology, Journal of Hematology and Oncology, Leukemia, Lancet Haematology, Blood, Blood Cancer Journal, and American Journal of Hematology.

As we used journal-based filtering, we used only the PubMed database for the search process. To better reflect the current state of reporting, we decided to include studies published during the previous 10 years (2012–2021). The complete search process is provided in Supplementary File 1. Our goal was to include at least 100 studies, and if that goal was not met after the initial screening, we were poised to include 5 additional years. The decision to aim for at least 100 studies was that there would be enough studies to be included to estimate the current reporting practices and quality.

Inclusion Criteria

We included all clinical or register-based studies conducted on humans, focused on leukemia and reporting survival, which is defined as a time-to-event outcome (including for example overall survival, event free survival). We included both retrospective and prospective studies. Furthermore, these studies should focus on pediatric patients (age 0–17 years) at the time of the intervention. As some studies consisted of both children and young adults, we included only those where over 50% of the participants were children.

Exclusion Criteria

Our exclusion criteria were as follows: all studies performed on animals; editorials and letters to the editor; reviews that did not present original data; and reports focusing on adults (over 50% of the study population being aged 18 or more years). Studies that did assess only survival of a single group without any comparator group(s) were excluded.

Search Process

The search was performed on January 4, 2022. The search result was uploaded to Covidence software. Two authors independently screened titles and abstracts. Throughout the screening process, conflicts were decided by mutual consensus or third-party opinion. All authors participated in this phase. Full texts were then assessed again by two authors independently.

Data Extraction and Outcome Measures

One author extracted the data to an Excel spreadsheet that was accepted by all authors. The most important extracted data are provided as Supplementary File 2, and all extracted data is available from the corresponding author. We extracted study years, number of patients, intervention, main outcome, crude survival methods used, and adjusted survival methods used. We defined the adequate use of survival methods as follows: crude survival method was reported and presented graphically; adjusted survival method was named; and, if the Cox PH model was used, the checking of the PH assumption and its result were reported. If the PH assumption was violated, we checked whether the authors stated how this was considered in the adjusted model and what modifications were made. If the authors did not comment on the background assumption, we labeled the PH assumption as possibly violated if the survival curves presented in the manuscript crossed clearly in the crude survival analysis. The crossing of the survival curves was assessed by three authors and labeled as (i) most likely PH violation, (ii) could cause PH violation, or (iii) no suspicion of PH violation. We sought mutual agreement for each of the suspected crossings. We focused on crossing of curves, as this is a cause or a sign of non-proportionality in the Cox PH model. We checked

supplementary files for additional information on statistical analyses if the authors mentioned supplements in the "Methods" section.

Permissions

Although we did not make a quantitative meta-analytic synthesis of the included papers, we performed our study process and data extraction systematically. Therefore, we have reported our manuscript according to the Preferred Reporting Item for Systematic Reviews and Meta-analyses (PRISMA) [17]. We provide the PRISMA statement in Supplementary File 3. Our study protocol was registered to PROSPERO (ID: CRD42022301378) and is available from https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022301378

Results

Our initial search retrieved 561 results and, after the screening of abstracts and full-text assessment, 103 studies were included in the analysis (Fig. 1). Most of the included studies were published in the Journal of Clinical Oncology (Table 1). The main interventions analyzed were genetics and treatment interventions. Most of the studies were prospective, and 31 were randomized controlled trials.

The most-applied crude survival method was Kaplan–Meier, as 96 (94%) of the 103 studies used it. Kaplan–Meier curves were presented graphically in 91 of the 96 studies. Of these 91 studies, only four (4%) presented 95% confidence intervals in the Kaplan–Meier graph. We noted crossing of survival curves in at least one figure in ten (10%) studies of those that presented

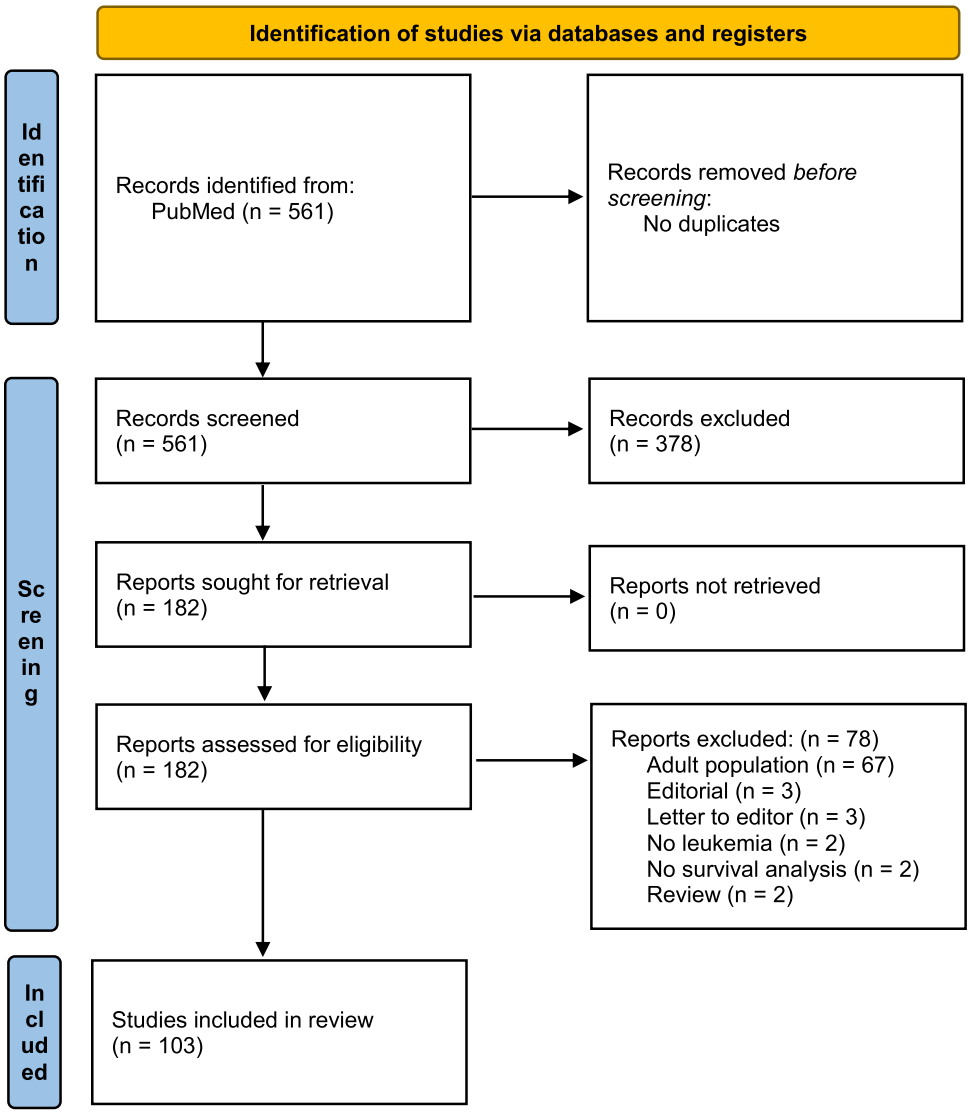

Fig. 1 Flow chart of the review process

Table 1 Characteristics of the included studies

	N	%
Journal		
American Journal of Hematolgy	10	10
Annals of Oncology	0	0
Blood	28	27
Blood Cancer Journal	3	3
JAMA Oncology	0	0
Journal of Clinical Oncology	37	36
Journal of Hematology and Oncology	1	1
Lancet Haematolgy	5	5
Lancet Oncolgy	7	7
Leukemia	12	12
Study design		
Prospective	77	75
Clinical	95	92
RCT	31	30
Intervention/Exposure		
Medicine or treatment	47	46
Gene	29	28
Laboratory parameters	19	18
Patient related factors	8	8

graphical survival curves. Of these ten studies, one was RCT and nine were observational studies. As only some individual studies used methods other than the Kaplan–Meier and Cox PH models, we focused on Cox PH model reporting. Adjusted survival analysis was performed in 80 (78%) of the included studies; in these, the Cox PH model was most used (77 studies, 96%). Of these 77 studies, 18 (23%) were RCT. Other adjusted methods used were the Fine and Gray, Pohar Perme, and odds-ratio methods.

The PH assumption was mentioned in 18 (23%) of the 77 studies that used the Cox PH model. Four (22%) of the RCTs and 14 (24%) of the observational studies. Only nine (12%) studies specifically stated how the assessment of the PH assumption was performed. Log-minus-log plots were used in three studies, log-time analysis in three studies, Schoenfeld residuals in two studies, and visual examination of survival curves in one study. Three studies stated that they had considered the PH assumption violation and used a time-dependent coefficient analysis for correction of the model.

We noted 10 studies with possible PH assumption violations and of these one was RCT. Two studies had clear violations, meaning that survival curves crossed clearly in Kaplan–Meier analysis, while eight studies most likely violated the PH assumption. In seven of these cases, the authors did not report or discuss the possibility of the non-proportionality of hazards.

Discussion

We examined the use and reporting of survival analysis methods in pediatric leukemia studies over the past 10 years. The most-applied crude survival analysis method was Kaplan–Meier, while the most-used adjusted survival method was Cox PH. The majority of the studies did not mention whether the PH background assumption was evaluated in the Cox PH model. An obvious or likely violation of the PH assumption was observed in 13% of the included studies. The reporting practices were similar regarding the PH assumption in RCTs and observational studies.

The reporting quality of Cox PH is not a new issue, as previous studies analyzing oncology and orthopedic studies have produced similar findings [9, 18–20]. However, this is to our knowledge the first study to examine the quality of survival method reporting in pediatric leukemia studies. Although we found few cases with obvious or likely crossing of survival curves or violation of the PH assumption, the vast majority of the studies did not report survival methods with the expected detail. To improve reporting and methodological quality, greater awareness is needed. For example, the use of statistical editors and reviewers by journals would improve reporting quality [21]. Likewise, reporting guidelines have been shown to improve and uniform reporting [22].

Reproducibility is a key part of science, meaning that studies should be presented in such a detailed manner that the methods could be repeated by a reader [23]. This is a rather ambitious goal. In statistical reporting, however, certain key elements are required for evaluation of the validity of results. For example, the basis of choices for statistical testing is important, and the Cox PH model is typically used under the PH assumption. In this scenario, the PH assumption should be checked, the testing methods named, and the handling of possible non-proportionality described. However, if authors do not present these details, readers are left wondering whether this was due to lack of knowledge or was an intentional decision made by the authors. One should bear in mind that in certain conditions, the Cox PH model may be applied regardless of the fulfillment of the PH assumption. Indeed, in reality, the expectation of continuously proportional hazards is biologically controversial [10, 24]. However, ignorance of the PH assumption should be an intentional and thoroughly justified choice with a careful consideration of the possibility of biased estimates due to the influence of time-varying covariate effects.

Violation of the proportionality issue often leads to biased estimates and therefore might result in misleading information regarding extremely important aspects, such as oncological treatments, and, in the worst-case scenario, to wrong treatment decisions. Instead of Cox PH, alternative methods have been suggested. For example, the restricted mean survival time method has less restrictive background assumptions, and interpretation of the produced estimates is rather clear [25, 26]. Interestingly, majority of the studies did not present confidence intervals in the Kaplan–Meier graphs, which makes it practically impossible to interpret the uncertainty of the presented findings.

This study examined the quality of reporting of survival analysis methods in pediatric leukemia studies. We acknowledge that some authors may have checked the PH assumption but not mentioned this in the manuscript. Hence, that a relatively low number of studies reported testing of the PH assumption does not necessarily mean that it was ignored. Another minor limitation of our study is that the evaluation of possible violations of the PH assumption was based on subjective visual examination of Kaplan–Meier survival curves instead of robust statistical testing, although visual examination is an excellent choice in some scenarios [27]. Therefore, we labeled probable cases of non-proportionality "likely" or "clear" to avoid overestimating problems, as for example, in cases with low number of participants, it is hard to interpret whether the curves are crossing or overlapping. Further limitation is that the most likely our results underestimate the issue of incorrect reporting as we have focused on the top five cited journals of oncology and hematology, and thus these issues may be more common in less cited journals. The main strengths of this study are that we did not have protocol deviations and employed a systematic approach to the evaluation of the statistical method reporting of the included studies.

Conclusions

Reporting of survival analysis methods in the majority of pediatric leukemia studies published in the highest-ranked oncology and hematology journals was inadequate. The Cox PH model was the most-applied adjusted survival analysis, but evaluation of the background assumptions was rarely reported. We also found a few cases where crossing survival curves, representing obvious violations of the PH assumption, were presented without justification or adjustment. Better statistical reporting is needed in pediatric leukemia studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42399-022-01367-y.

Author Contribution Ilari Kuitunen: original idea, data gathering, protocol writing, screening of the abstracts and texts, data extraction, analyses, writing of the first draft.

Atte Nikkilä: original idea, data gathering, screening of the abstracts and texts, data extraction, analyses, revisions and commenting of the manuscript.

Ville Ponkilainen: data gathering, screening of the abstracts and texts, data extraction, revisions and commenting of the manuscript.

Mikko Uimonen: data gathering, screening of the abstracts and texts, data extraction, revisions and commenting of the manuscript.

Olli Lohi: original idea, supervision, resources, data extraction, revisions and commenting of the manuscript.

Funding Open access funding provided by University of Eastern Finland (UEF) including Kuopio University Hospital.

Data Availability Available as supplement.

Code Availability Not applicable.

Declarations

Ethics Approval Not applicable in systematic review.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Conflict of Interest The authors declare no competing interests.

References

1. Gyawali B, Tessema FA, Jung EH, Kesselheim AS. Assessing the justification, funding, success, and survival outcomes of randomized noninferiority trials of cancer drugs: a systematic review and pooled analysis. JAMA Netw Open. 2019;2(8): e199570. https://doi.org/10.1001/jamanetworkopen.2019.9570.
2. Bell Gorrod H, Kearns B, Stevens J, et al. A review of survival analysis methods used in NICE technology appraisals of cancer treatments: consistency, limitations, and areas for improvement. Med Decis Making. 2019;39(8):899–909. https://doi.org/10.1177/0272989X19881967.
3. Porter ME. What is value in health care? N Engl J Med. 2010;363(26):2477–81. https://doi.org/10.1056/NEJMp1011024.
4. 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.
5. George B, Seals S, Aban I. Survival analysis and regression models. J Nucl Cardiol. 2014;21(4):686–94. https://doi.org/10.1007/s12350-014-9908-2.
6. Cox DR. Regression models and life-tables. J Roy Stat Soc. Series B (Methodological). 1972;34(2):187–220.
7. Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CGM. Time-varying covariates and coefficients in Cox regression models. Ann Transl Med. 2018;6(7):121. https://doi.org/10.21037/atm.2018.02.12.
8. Horiguchi M, Hassett MJ, Uno H. How do the accrual pattern and follow-up duration affect the hazard ratio estimate when the proportional hazards assumption is violated? Oncologist. 2019;24(7):867–71. https://doi.org/10.1634/theoncologist.2018-0141.
9. Kuitunen I, Ponkilainen VT, Uimonen MM, Eskelinen A, Reito A. Testing the proportional hazards assumption in cox regression and dealing with possible non-proportionality in total joint arthroplasty research: methodological perspectives and review. BMC Musculoskelet Disord. 2021;22(1):489. https://doi.org/10.1186/s12891-021-04379-2.
10. Stensrud MJ, Hernán MA. Why test for proportional hazards? JAMA. 2020;323(14):1401–2. https://doi.org/10.1001/jama.2020.1267.
11. Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol. 2013;13:152. https://doi.org/10.1186/1471-2288-13-152.
12. 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.
13. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med. 2010;152(11):726–32. https://doi.org/10.7326/0003-4819-152-11-201006010-00232.
14. 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. https://doi.org/10.1016/j.ijsu.2014.07.013.
15. Turkiewicz A, Luta G, Hughes HV, Ranstam J. Statistical mistakes and how to avoid them - lessons learned from the reproducibility crisis. Osteoarthr Cartil. 2018;26(11):1409–11. https://doi.org/10.1016/j.joca.2018.07.017.
16. Block JA. The reproducibility crisis and statistical review of clinical and translational studies. Osteoarthritis Cartilage. 2021;29(7):937–8. https://doi.org/10.1016/j.joca.2021.04.008.
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372: n71. https://doi.org/10.1136/bmj.n71.
18. 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. https://doi.org/10.1186/s12878-016-0055-7.
19. 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). https://doi.org/10.1371/journal.pone.0154870.
20. 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. Medicine (Baltimore). 2017;96(50):e9204. https://doi.org/10.1097/MD.0000000000009204.
21. Hardwicke TE, Goodman SN. How often do leading biomedical journals use statistical experts to evaluate statistical methods? The results of a survey. PLoS ONE. 2020;15(10): e0239598. https://doi.org/10.1371/journal.pone.0239598.
22. Vandenbroucke JP. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK… and CONSORT: for whom does the guideline toll? J Clin Epidemiol. 2009;62(6):594–6. https://doi.org/10.1016/j.jclinepi.2008.12.003.
23. Gamble C, Krishan A, Stocken D, et al. Guidelines for the content of statistical analysis plans in clinical trials. JAMA. 2017;318(23):2337–43. https://doi.org/10.1001/jama.2017.18556.
24. Uno H, Tian L. Is the log-rank and hazard ratio test/estimation the best approach for primary analysis for all trials? JCO. 2020;38(17):2000–1. https://doi.org/10.1200/JCO.19.03097.
25. O'Quigley J. Testing for differences in survival when treatment effects are persistent, decaying, or delayed. JCO. 2022:JCO.21.01811. https://doi.org/10.1200/JCO.21.01811.
26. Freidlin B, Korn EL. Methods for Accommodating nonproportional hazards in clinical trials: ready for the primary analysis? JCO. 2019;37(35):3455–9. https://doi.org/10.1200/JCO.19.01681.
27. Persson I, Khamis HJ. A comparison of graphical methods for assessing the proportional hazards assumptions in the Cox model. J Stat Appl. 2007;2 (1–4), 1–32. https://corescholar.libraries.wright.edu/math/271.