Aalborg Universitet

Detecting deviations from the efficacy and safety results of single-arm trials using real-world data

The case of a CAR-T cell therapy in B-cell lymphoma

Jakobsen, Lasse Hjort; Callréus, Torbjörn; Sessa, Maurizio; Jerkeman, Mats; Andersen, Morten; El-Galaly, Tarec Christoffer

Published in:
Pharmacoepidemiology and Drug Safety

DOI (link to publication from Publisher):
10.1002/pds.5195

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Jakobsen, L. H., Callréus, T., Sessa, M., Jerkeman, M., Andersen, M., & El-Galaly, T. C. (2021). Detecting deviations from the efficacy and safety results of single-arm trials using real-world data: The case of a CAR-T cell therapy in B-cell lymphoma. *Pharmacoepidemiology and Drug Safety*, *30*(4), 514-519. https://doi.org/10.1002/pds.5195

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: December 04, 2025

Check for updates

Detecting deviations from the efficacy and safety results of single-arm trials using real-world data: the case of a CAR-T cell therapy in B-cell lymphoma

Short running title: Monitoring of CAR-T cell therapy

Lasse Hjort Jakobsen[1,2], Torbjörn Callréus[3*], Maurizio Sessa[3], Mats Jerkeman[4], Morten Andersen[3], Tarec Christoffer El-Galaly [1,2]

1. Department of Hematology, Aalborg University Hospital, Aalborg, Denmark
2. Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
3. Pharmacovigilance Research Center, Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark
4. Department of Oncology, Lund University, Lund, Sweden.

*Present affiliation: Malta Medicines Authority, San Ġwann, Malta

Corresponding author
Lasse Hjort Jakobsen, PhD
Department of Hematology
Aalborg University Hospital
Sdr. Skovvej 15, Aalborg DK-9000, Denmark
Email: lasse.j@rn.dk or lahja@dcm.aau.dk

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/pds.5195

This article is protected by copyright. All rights reserved.

Accepted Article

Key Words: CAR-T cell therapy, monitoring, single-arm trial

Take home messages:

- An increasing number of marketing authorizations are based on single-arm trials, which increases the demand for better post-authorization monitoring strategies.
- In a general population of 5 million, the predicted number of relapsed-refractory large B-cell lymphoma patients treated with chimeric antigen receptor T cell therapy was 66.
- Assuming a 10% true absolute decrease in 1-year OS among real-world patients compared to trial participants, 10.5 years of data accrual are needed to achieve 80% power for detecting a significant decrease.
- Power calculations can provide insights into when real-world data sources can realistically detect significant deviations from the safety and efficacy established in trials.
- For therapeutics administered infrequently, such as chimeric antigen receptor T cell therapy, combining data from multiple countries with similar health care systems can accelerate the confirmation of efficacy and safety.

Word count excluding abstract, tables, figures and references: 2083

Parts of the content of this report were presented by poster at the 2019 International Conference on Pharmacoepidemiology & Therapeutic Risk Management (Philadelphia, USA).

This article is protected by copyright. All rights reserved.

Accepted Article

Abstract

Purpose: Personalized therapies are leading to an increasing number of marketing authorizations based on single-arm trials, which increases the demand for better post-authorization monitoring strategies. The aim of the present study was to estimate the power over time as data accrue in population-based registries for detecting deviations from the expected efficacy/safety of chimeric antigen receptor T cell (CAR-T) therapy approved for relapsed/refractory large B-cell lymphoma (RR-LBCL).

Methods: The number of real-world RR-LBCL patients was projected over time in a general population of 5, 15, and 25 million citizens using lymphoma registry data. For each scenario, we computed the power over time for detecting significant deviations in efficacy (1-year overall survival [1yOS]) when comparing to historical controls (SCHOLAR-1 study; 1yOS, 28%) and RR-LBCL patients treated with CAR-T cell therapy in a single-arm trial (ZUMA-1; 1yOS, 59%) as well as deviations in selected adverse events (grade ≥3 aphasia) from the ZUMA-1 trial. We assumed a 10% absolute deviation in 1yOS (efficacy) and a relative increase of 50% in grade ≥3 aphasia (safety).

Results: Assuming a general population of 5, 15, and 25 million, the accrual time needed to achieve 80% power for detecting a significant increase over the 1yOS reported in SCHOLAR-1 was 9, 4, and 3 years, respectively, while 80% power for detecting a significant decrease in 1yOS compared to ZUMA-1 required 10.5, 4.5, and 3 years of data accrual, respectively. However, corresponding estimates for aphasia were >20, 8, and 5 years, respectively.

Conclusions: Projections of the statistical power for detecting important deviations in efficacy/safety from that reported in pivotal clinical trials(s) provide critical information about the expected performance of post-authorization monitoring programs.

This article is protected by copyright. All rights reserved.

Accepted Article

Introduction

As novel cancer therapies are increasingly targeting small subsets of patients who fail to respond to broadly used first line therapies, marketing authorizations are frequently granted based on single-arm trials.[1] In the absence of randomized controlled trials, new approaches for evidence generation are important.[2–4] Alternative post-marketing evidence generation may be particularly relevant for treatments of serious diseases such as cancers when the treatment effect in the single-arm trial is substantially better than expected from historical data. In these situations, randomized trials become unethical and with little incentive for patients to participate. Without randomized data, the post-authorization monitoring plan becomes critical to confirm the risk/benefit balance of novel therapies approved based on single-arm trials. Potential discrepancies in the risk/benefit outcomes between the trial and real-world subjects should be monitored. Differences should be explored as they could suggest limitations concerning external validity of the registrational trial, which may originate from patient selection or important variations in clinical practice.

The approval of the chimeric antigen receptor T (CAR-T) cell therapy, axicabtagene ciloleucel (axi-cel), for relapsed/refractory large B-cell lymphoma (RR-LBCL) was based on ZUMA-1, a single-arm trial.[5] Enrolling 108 RR-LBCL patients for whom effective standard therapies were not available and reporting a 1-year overall survival (OS) of 59% paved the way for accelerated approval without regulatory requirement of a confirmatory phase 3 trial, although the marketing authorization holder is obligated to conduct a non-interventional post-authorization safety study.[6] To ensure that the results of this novel cellular therapy can be generalized to patients treated in the real-world setting, post-authorization monitoring programs can be used to detect early deviations in efficacy and safety from the ZUMA-1 results as well as confirm the expected outcome improvement over historical patients.

This study aimed to demonstrate the statistical power over time for detecting significant deviations in selected efficacy and safety outcomes in real-world patients treated with axi-cel compared to historical controls and patients treated in the pivotal ZUMA-1 trial. In this example, the calculations are based on data accrual in the Nordic population-based registries.[7]

Methods

The ability to detect important deviations from findings in single-arm trials can be quantified by statistical power, i.e., the probability of being able to identify a true deviation. Power calculations are directly impacted by the expected number of patients available for analysis and the effect size of the expected deviation. The estimated annual rate of RR-LBCL patients treated with CAR-T cell therapy derived from incidences and treatment outcomes from registry and trial data was 3.3/1,000,000 person-years (Figure S1).[8–11] The annual number of patients receiving CAR-T cell therapy was estimated by multiplying the annual rate and the size of the general population for which we considered three scenarios: 1) 5 million (approximate size of Denmark), 2) 15 million (approximate combined size of Denmark and Sweden), and 3) 25 million (approximate combined size of Denmark, Sweden, Norway, and Finland).

This article is protected by copyright. All rights reserved.

Accepted Article

First, the power to detect significant improvements in efficacy compared to historical controls was determined using the international SCHOLAR-1 study of 636 RR-LBCL patients.[12] The SCHOLAR-1 study was designed to provide a relevant efficacy benchmark for ZUMA-1 and so included patients from two observational cohorts and two phase 3 trials, who were treated for RR-LBCL with therapies available at the time (prior to the approval of axi-cel). The SCHOLAR-1 study reported a 1-year OS of 28%. The power to detect a significant improvement in efficacy (1-year OS) compared to the SCHOLAR-1 results was computed assuming a true 15%, 10%, and 5% absolute increase in 1-year OS among real-world patients receiving axi-cel.

Secondly, we computed the power to detect a significantly lower efficacy than reported in the ZUMA-1 trial (1-year OS, 59%), assuming a true 15%, 10%, and 5% absolute decrease in 1-year OS.[5]

Lastly, the power to detect deviations from the ZUMA-1 trial in terms of safety, here exemplified by the incidence of grade ≥3 aphasia (ZUMA-1 incidence, 7%), was computed under the assumption of a 50% and 25% relative increase in the incidence of grade ≥3 aphasia.

The type I error rate was set to 5%. The one-sided one-sample formula for statistical power in the context of binary outcomes described by Fleiss *et al.* was used for computing the power under the three general population scenarios (see Supplementary for details).[13] The power was estimated in half-year intervals starting from January 1st 2021 based on the predicted number of patients accrued in each interval in health databases with 100% population coverage.

Results

With a general population size of 5, 15, and 25 million, the predicted numbers of accrued RR-LBCL patients who received CAR-T cell therapy were 66, 198, and 330 by January 2025, respectively, and 148, 446, and 743 by January 2030, respectively.

Assuming a true absolute improvement of 10% in 1-year OS and a general population of 5 million, the time to reach 80% power to detect a significant deviation from the SCHOLAR-1 results was 9.0 years (Figure 1). By increasing the general population to 15 or 25 million, the time was reduced to 4.0 and 3.0 years, respectively. When the true absolute improvement was 5%, the corresponding estimates with a general population of 5, 15, and 25 million were >20, 11.5, and 7.5 years, respectively. Assuming a true absolute improvement of 15%, the corresponding estimates were 5.0, 2.5, and 2.0 years.

Assuming a true absolute decrease in 1-year OS of 10% compared to ZUMA-1, the time to achieve 80% power was 10.5, 4.5, and 3.0 years for a general population of 5, 15, and 25 million, respectively. Corresponding estimates when assuming a true absolute decrease of 5% were >20, 13.5, and 8.5 years. Assuming a true absolute decrease of 15%, the time needed to achieve 80% power was 5.5, 2.5, and 2.0 years for a general population of 5, 15, and 25 million, respectively.

When assuming a 50% relative increase in grade ≥3 aphasia (i.e., an absolute increase from 7% to 10.5%), the time to reach a power of 80% for detecting a significant deviation in safety was >20 years in a general population of 5 million which was reduced to 8.0 and 5.0 years for a general population of 15 and 25 million, respectively (Figure 2). Under the assumption of a

This article is protected by copyright. All rights reserved.

Accepted Article

true 25% (i.e., an absolute increase from 7% to 8.75%) increase in grade ≥3 aphasia, the time to achieve 80% power was 17.5 years for a general population of 25 million.

Discussion

Several years of patient accrual in public health care databases are typically required to obtain meaningful sample sizes for analysis of drug effectiveness and safety. To accelerate patient accrual, international collaborations involving countries with similar data infrastructure are needed. The Nordic countries share many similarities in terms of infrastructure such as fully publicly funded health care systems and a tradition for national health care registries with high coverage and quality.[7] This makes Nordic data collaborations appealing for monitoring efficacy and safety of novel therapies in routine clinical practice. In this study, we considered the novel CAR-T cell therapy as this represents a paradigm shift in the treatment of RR-LBCL and is the first therapy based on genetically modified T-cells to become available for routine use.[14] CAR-T cell therapies have shown impressively high efficacy in patients with very dismal outcomes on traditional therapies, but the price of CAR-T cell therapy, including supportive care and setup needed to provide the treatments safely, could exceed 500,000USD per treatment.[15] Additionally, the complex production of these genetically modified T-cells could result in some variability in the quality of the product, potentially affecting effectiveness and safety.[16] Therefore, early planning of treatment monitoring (preferably before market introduction of the new treatments) is critical to ensure that effectiveness and safety meet the expectations and provide reasonable value to the society. Efficacy and safety could also be confirmed in subsequent clinical trials, possibly also investigating the efficacy in earlier treatment lines, which may reduce the relevance of real-world data once the trial data are mature. However, even in cases where subsequent clinical trials are initiated, real-world data remain important since knowledge on significant efficacy/safety deviations would be important for regulatory approval strategies for novel drugs in the future.

In a recent study of 275 patients treated with axi-cel in a standard of care setting, efficacy and safety were comparable to that reported in ZUMA-1, suggesting feasibility of axi-cel outside clinical trials.[17] However, cellular therapies involving genetic modifications like CAR-T cell therapy require a setup where signals of serious late effects years after therapy are captured effectively.

Importantly, with a general population of 5 million, many years of data accrual are needed to achieve 80% power to detect significant deviations from the SCHOLAR-1 and ZUMA-1 trials in terms of efficacy and safety. For all considered scenarios and endpoints, the time to reach 80% power was reduced by ≥2.5 years by increasing the general population size from 5 to 15 million, but the reduction was limited when increasing the general population from 15 to 25 million in cases with a true absolute deviation of ≥10%, and therefore the added complexity of managing additional data should be considered. In our calculations, the time needed to detect deviations in terms of safety was generally longer compared to detecting deviations in terms of efficacy, even when assuming a 50% increase. Hence, for early detection of deviations from ZUMA-1 in terms of safety (grade ≥3 aphasia), extensive collaboration on data collection is warranted.

This article is protected by copyright. All rights reserved.

Accepted Article

The small size of the ZUMA-1 trial will limit the precision of any comparison between trial patients and those treated in clinical practice in the post-marketing setting. Moreover, in our calculations we assumed that patients treated in routine clinical practice share the same characteristics as patients enrolled in ZUMA-1/SCHOLAR-1. As frail and elderly patients typically are underrepresented in clinical trials, this will likely not be the case. The real-world population should be characterized and compared to the trial population, and statistical techniques, such as inverse probability of treatment weighting, matching, and standardization, may be needed to adjust for imbalances. However, this may not only be an issue of adjustment. Some real-world patients receiving CAR-T cell therapy may differ considerably from those in the trial population, e.g., due to comorbidities associated with a high risk of adverse outcomes, and the use of CAR-T cell therapy in these patients may not be warranted based on the ZUMA-1 trial. Another limitation is the assumption of a single statistical test at one point in time when sequential monitoring may be a more relevant approach. However, sequential monitoring increases the likelihood of false positives and therefore requires the use of techniques, such as spending functions, to adjust the false positive rate.[18]
In this study, we only considered the power to detect a significant deviation in efficacy and safety outcomes of accrued real-world data from the pivotal trial results. For regulatory decisions, the absolute incidence of an adverse event in itself may in some cases be important even though a significant difference between real-world patients and trial participants has not been observed, especially in the case of rare and serious events. Finally, when there are more than one drug marketed for the same indication, comparisons between these drugs may be more relevant than using a pivotal trial as reference.

In conclusion, when post-authorization monitoring programs use external controls from a clinical trial, projections of the expected statistical power provide a time frame for monitoring drug efficacy and safety, thereby facilitating pharmacovigilance planning.[19] For therapeutics administered infrequently, such as CAR-T cell therapy, combining health care databases from multiple countries facilitates rapid monitoring of efficacy and safety.

Funding

Morten Andersen and Maurizio Sessa belongs to the Pharmacovigilance Research Center, Department of Drug Design and Pharmacology, University of Copenhagen, funded by the Novo Nordisk Foundation (NNF15SA0018404). No funding was specifically received for the preparation of this article. Torbjörn Callréus contributed to this report while affiliated with the Center. Lasse Hjort Jakobsen received funding by Aalborg University.

Conflict of Interest

This article is protected by copyright. All rights reserved.

Accepted Article

Mats Jerkeman has received research grants from Gilead, Abbvie, Janssen and Celgene, and personal fees from Gilead, Janssen, Novartis, Acerta, Roche and Celgene. Tarec Christoffer El-Galaly is employed by F. Hoffmann-La Roche, Ltd, Basel. Morten Andersen has participated in research projects funded by AstraZeneca, H. Lundbeck & Mertz, Novartis, Pfizer and Janssen, with grants paid to the institutions where he has been employed, and has personally received fees from Atrium, the Danish Pharmaceutical Industry Association, for leading and teaching pharmacoepidemiology courses. Lasse Hjort Jakobsen, Torbjörn Callréus, and Maurizio Sessa have no conflicts of interest that are directly relevant to the content of this article.

ORCID

Lasse Hjort Jakobsen	0000-0001-6575-7686
Torbjörn Callréus	0000-0001-9903-5817
Maurizio Sessa	0000-0003-0874-4744
Mats Jerkeman	0000-0003-4509-6707
Morten Andersen	0000-0001-7029-2860
Tarec Christoffer El-Galaly	0000-0002-4406-380X

References

1. Anderson M, Naci H, Morrison D, Osipenko L, Mossialos E. A review of NICE appraisals of pharmaceuticals 2000–2016 found variation in establishing comparative clinical effectiveness. *J Clin Epidemiol* 2019; **105**: 50–59. doi:10.1016/j.jclinepi.2018.09.003.
2. Ngo L, Hee S-W, Lim L-C, *et al.* Prognostic factors in patients with diffuse large B cell lymphoma: Before and after the introduction of rituximab. *Leuk Lymphoma* 2008; **49**: 462–469.
3. Carrigan G, Whipple S, Capra WB, *et al.* Using Electronic Health Records to Derive Control Arms for Early Phase Single-Arm Lung Cancer Trials: Proof-of-Concept in Randomized Controlled Trials. *Clin Pharmacol Ther* 2020; **107**: 369–377. doi:10.1002/cpt.1586.
4. Eichler HG, Koenig F, Arlett P, *et al.* Are Novel, Nonrandomized Analytic Methods Fit for Decision Making? The Need for Prospective, Controlled, and Transparent Validation. *Clin Pharmacol Ther* 2020; **107**: 773–779. doi:10.1002/cpt.1638.
5. Neelapu SS, Locke FL, Bartlett NL, *et al.* Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N Engl J Med* 2017; **377**: 2531–2544. doi:10.1056/NEJMoa1707447.
6. Yescarta : EPAR - Product Information. *Eur Med Agency* 2018. Available at: https://www.ema.europa.eu/en/documents/product-information/yescarta-epar-product-information_en.pdf. Accessed June 11, 2020.

This article is protected by copyright. All rights reserved.

Accepted Article

7. Callréus T, El-Galaly TC, Jerkeman M, de Nully Brown P, Andersen M. Monitoring CAR-T-Cell Therapies Using the Nordic Healthcare Databases. *Pharmaceut Med* 2019; **33**: 83–88. doi:10.1007/s40290-019-00270-7.
8. Li Y, Wang Y, Wang Z, Yi D, Ma S. Racial differences in three major NHL subtypes: Descriptive epidemiology. *Cancer Epidemiol* 2015; **39**: 8–13. doi:10.1016/j.canep.2014.12.001.
9. Ekberg S, Jerkeman M, Andersson P-O, *et al.* Long-term survival and loss in expectancy of life in a population-based cohort of 7114 patients with diffuse large B-cell lymphoma. *Am J Hematol* 2018; **93**: 1020–1028. Available at: http://doi.wiley.com/10.1002/ajh.25147. Accessed July 24, 2018.
10. Jakobsen LH, Bøgsted M, Brown P de N, *et al.* Minimal Loss of Lifetime for Patients With Diffuse Large B-Cell Lymphoma in Remission and Event Free 24 Months After Treatment: A Danish Population-Based Study. *J Clin Oncol* 2017; **35**: 778–784. doi:10.1200/JCO.2016.70.0765.
11. Gisselbrecht C, Glass B, Mounier N, *et al.* Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol* 2010; **28**: 4184–4190. doi:10.1200/JCO.2010.28.1618.
12. Crump M, Neelapu SS, Farooq U, *et al.* Outcomes in refractory diffuse large B-cell lymphoma: Results from the international SCHOLAR-1 study. *Blood* 2017; **130**: 1800–1808. doi:10.1182/blood-2017-03-769620.
13. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3td ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2003. doi:10.1002/0471445428.
14. El-Galaly TC, Cheah CY, Kristensen D, *et al.* Potentials, challenges and future of chimeric antigen receptor T-cell therapy in non-Hodgkin lymphomas. *Acta Oncol (Madr)* 2020. doi:10.1080/0284186X.2020.1741680.
15. Whittington MD, McQueen RB, Ollendorf DA, *et al.* Long-term Survival and Cost-effectiveness Associated With Axicabtagene Ciloleucel vs Chemotherapy for Treatment of B-Cell Lymphoma. *JAMA Netw open* 2019; **2**: e190035. doi:10.1001/jamanetworkopen.2019.0035.
16. Schwab RD, Bedoya DM, King TR, Levine BL, Posey AD. Approaches of T Cell Activation and Differentiation for CAR-T Cell Therapies. In: *Methods in Molecular Biology*.Vol 2086. Humana Press Inc., 2020; 203–211. doi:10.1007/978-1-0716-0146-4_15.
17. Nastoupil LJ, Jain MD, Feng L, *et al.* Standard-of-Care Axicabtagene Ciloleucel for Relapsed or Refractory Large B-Cell Lymphoma: Results From the US Lymphoma CAR T Consortium. *J Clin Oncol* 2020: JCO.19.02104. doi:10.1200/jco.19.02104.
18. Proschan MA, Lan KKG, Wittes JT. *Statistical Monitoring of Clinical Trials: A Unified Approach*. Springer, 2006. doi:10.1007/978-0-387-44970-8.
19. ICH Harmonised Guideline on Pharmacovigilance Planning (E2E). 2004. Available at: https://database.ich.org/sites/default/files/E2E_Guideline.pdf. Accessed May 28, 2020.

This article is protected by copyright. All rights reserved.

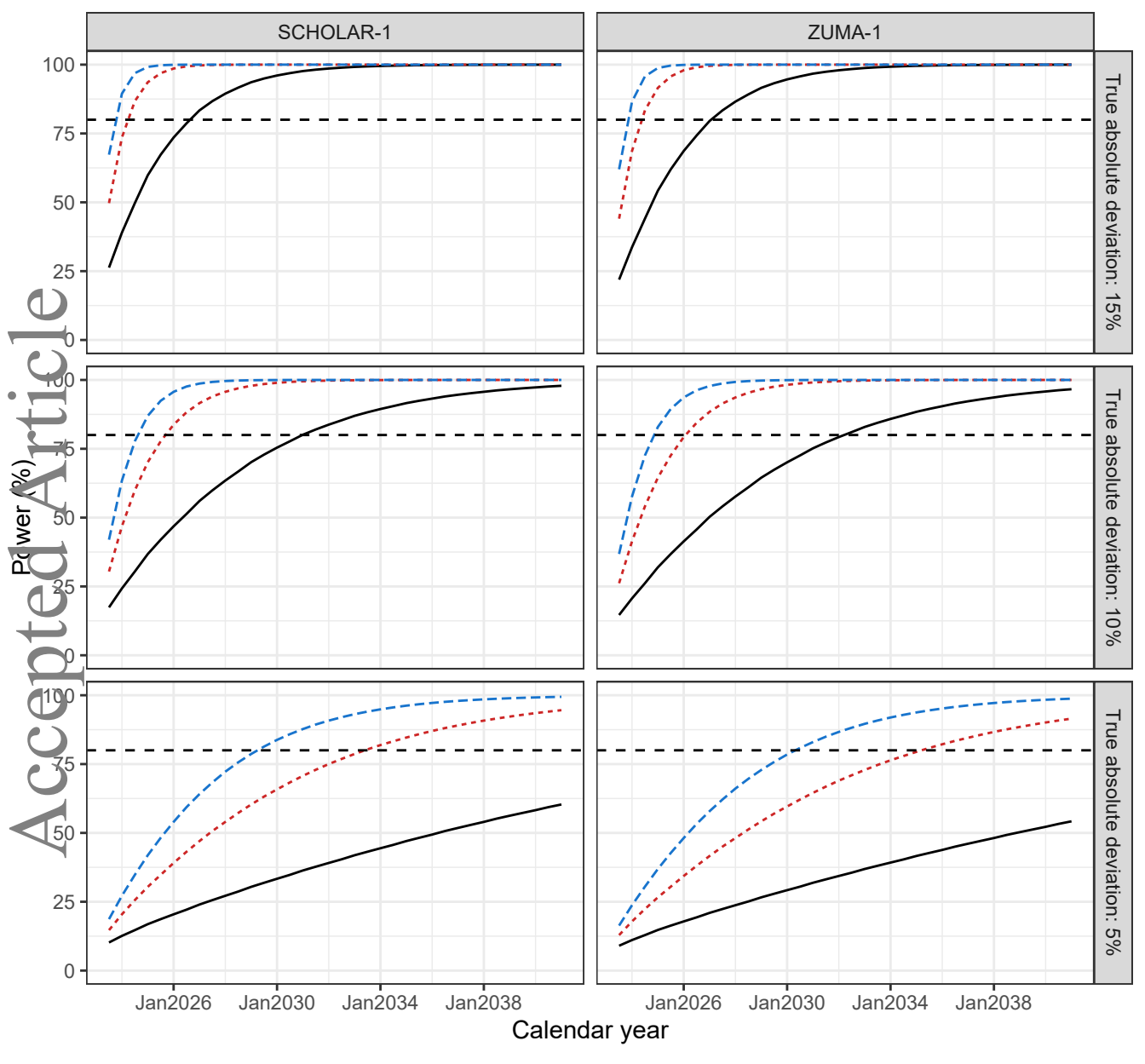
Figures

Figure 1: The estimated statistical power over time for detecting important deviations from the SCHOLAR-1 study and ZUMA-1 trial in terms of efficacy based on the projected number of relapsed/refractory large B-cell lymphoma patients in general populations of 5, 10, and 15 million citizens. The power was calculated for 1-year OS (SCHOLAR-1, 28%; ZUMA-1, 59%) under the assumption of a true absolute deviation of 15%, 10%, and 5%. The horizontal line indicates a statistical power of 80%.

Figure 2: The estimated statistical power over time for detecting important deviations from the the ZUMA-1 trial in terms of safety based on the projected number of relapsed-refractory large B-cell lymphoma patients in general populations of 5, 10, and 15 million citizens. The power was calculated for therapy-related grade ≥3 aphasia (ZUMA-1 proportion: 7%) assuming a true deviation of 50% and 25%. The horizontal line indicates a statistical power of 80%.

Accepted Article

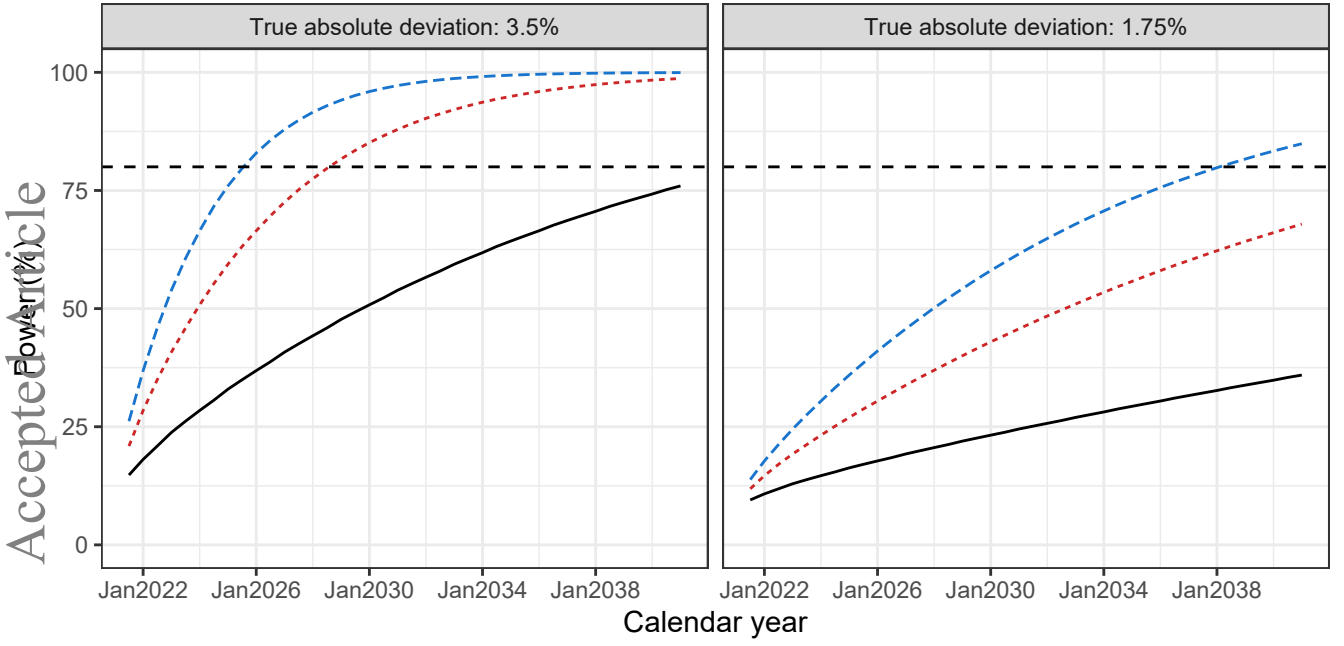
This article is protected by copyright. All rights reserved.



Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article



This article is protected by copyright. All rights reserved.

Population size — 5 million ··· 15 million - - 25 million