

# Learning Methodological Lessons from Exemplar Studies in Nephrology: PEXIVAS and Sample Size Calculation

Vicki Sandys <sup>1</sup> and Donal J. Sexton <sup>2,3</sup>

KIDNEY360 3: 1427–1430, 2022. doi: <https://doi.org/10.34067/KID.0005052021>

## Key Learning Points

- Sample size calculations are fundamental to the power of a study. Estimates of the necessary sample size to achieve a given treatment effect size can be performed pre-trial based on available literature.
- Sample size calculations take into account the chosen effect size i.e. the minimal effect of the treatment which would be considered clinically relevant. Smaller effect sizes require a sufficiently large sample size or higher event rate to detect differences between groups. Larger effect sizes are sometimes chosen for study in order to justify the risks of adverse events and expense of certain treatments.
- Lower than expected event rates, mediated by censoring and competing risks, period effects, and inclusion of patients at lower risk of the outcome, can impact the power of the study.
- Time-to-event study designs adjust for censoring and competing risks.
- If a lower than expected event rate is highlighted during an interim analysis, strategies include increasing the sample size, either by changing the inclusion or exclusion criteria, increasing the number of study centers or increasing the study duration.

## Introduction

Calculating an appropriate sample size is vital to ensure a clinical trial has adequate statistical power to detect significant differences between groups. Overestimating sample size can result in infeasible trials. Underestimating sample sizes increases the risk of missing small to moderate treatment effects (1,2). Sample size calculation depends upon several factors, including the prespecified significance level and desired degree of confidence (power of the study), the variability of the response to treatment, the probability of censoring and competing risks during the trial, and the estimated treatment effect size (1,2). These estimates can be inferred before the study based on the available literature. However, in certain instances, the assumptions on which the sample size was calculated may be insufficient, such as when high-quality studies to inform study design are lacking or if changes in population survival occur over time (1). In this short education piece, we examine the effect of sample size estimates using PEXIVAS as an exemplar study.

Plasma exchange and glucocorticoid dosing in the treatment of antineutrophil cytoplasm antibody-associated vasculitis study (PEXIVAS) was a landmark randomized trial in the specialty of nephrology and provides the best available evidence to date on induction

treatment for ANCA vasculitis with plasma exchange (PLEX) and corticosteroids (3). Before the publication of PEXIVAS, the use of PLEX and higher-dose corticosteroids in ANCA vasculitis was widespread in induction protocols (3,4). The results of PEXIVAS demonstrated no benefit of PLEX on the primary composite outcome of death and ESRD (hazard ratio 0.86, 95% confidence interval, 0.65 to 1.13) after a median follow-up of 2.9 years and called into question the benefits of PLEX in the treatment induction of ANCA vasculitis. The design of the study employed several nuanced methodological techniques, which the nephrology audience may benefit from exploring, including an adaptive adjustment for sample size during the trial following interim analysis (1,2,5).

The first learning point for general readers is to highlight the time component of the sample size calculation in time-to-event studies. Without a time-to-event component to the primary outcome, the sample size calculation is similar to that of studies assessing a difference in proportion between study groups effected at a fixed time point (2). However, the addition of a time-to-event analysis requires dealing with the issue of censored data and competing events, while full compliance is assumed. Censoring essentially denotes the handling of incomplete data arising

<sup>1</sup>Royal College of Surgeons in Ireland, Dublin, Ireland

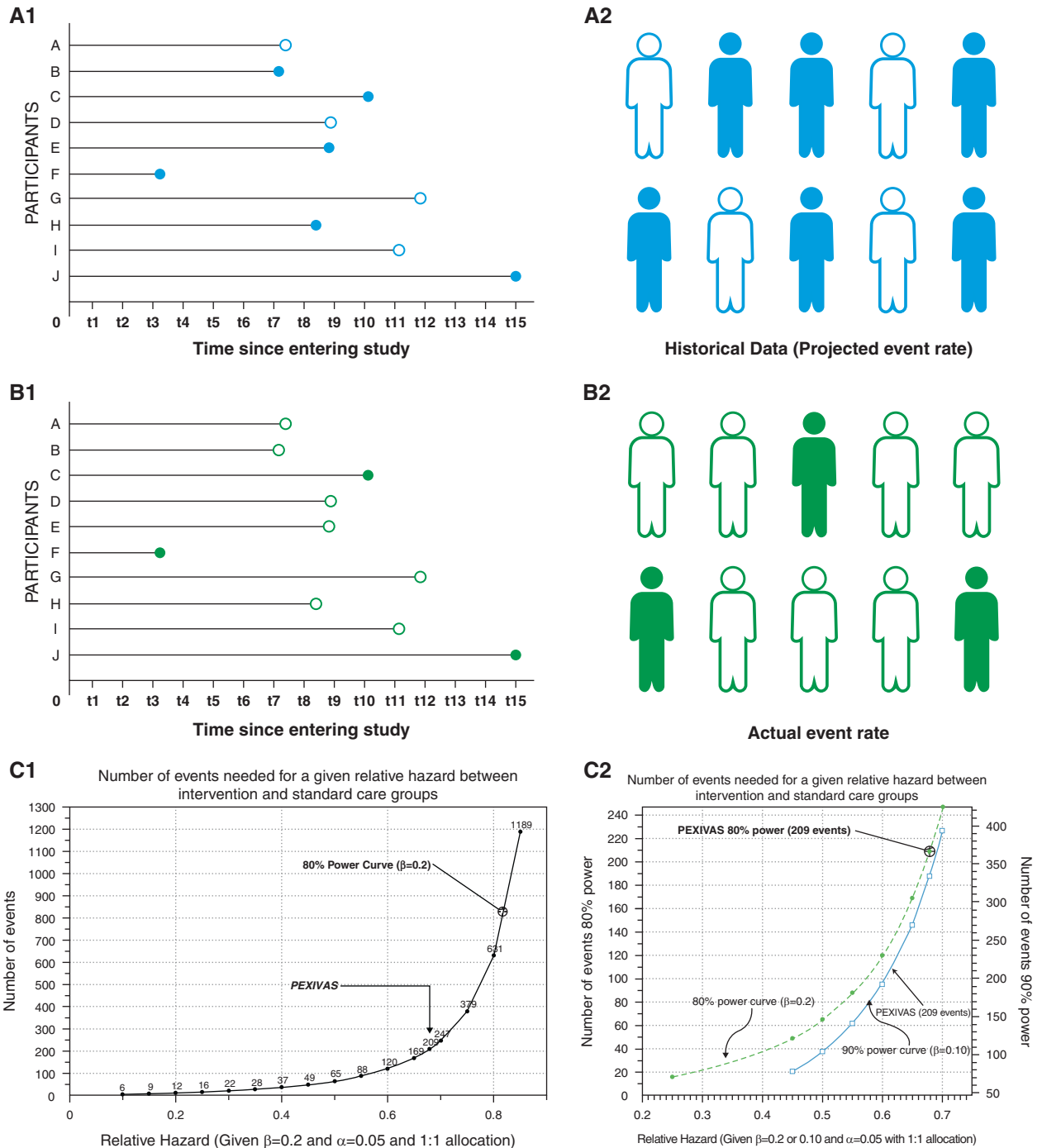
<sup>2</sup>St. James's Hospital, Dublin, Ireland

<sup>3</sup>School of Medicine, Trinity College Dublin, Dublin, Ireland

**Correspondence:** Prof. Donal J. Sexton, Renal Unit, St. James's Hospital, Dublin 8 D08 NHY1, Ireland. Email: [dosexton@tcd.ie](mailto:dosexton@tcd.ie)

due to different possibilities over follow-up, including experiencing the primary outcome, living without experiencing the event of interest, or no longer participating in the study by end of follow-up (5).

Time-to-event designs adjust for this type of data. The effective power of a study to detect a threshold level of treatment effect size can depend on the event rate seen (1). However, one of the interesting issues that can arise if



**Figure 1. | The potential impact of censoring on the event rate in randomized trials and PEXIVAS.** A hypothetical situation whereby projected event rates for a new study with a time to event primary outcome (B1 and B2) are based on event rates from a historical study (A1 and A2) with participants from the same study population (e.g., IgA nephropathy). One can appreciate how an event rate lower than expected (3/10 in [B2] versus 6/10 in [A2]) leads to more censored data over the duration of the study. Each line represents one participant. Solid dots and figures are events, and hollow dots and figures are censored data. t1–t15 represents time from first follow-up visit to the last in a hypothetical study. (C1 and C2) In PEXIVAS, the event rate corresponds to 80% power to detect a relative hazard ratio of 0.68, which is very close to what the investigators had planned. To the right of the PEXIVAS event rate, the reader can appreciate that in order to exclude a smaller treatment effect of PLEX, more events would likely be needed. *B*, probability of type II error.

censoring occurs more frequently than expected (Figure 1) is that the event rate in a study is lower than expected. If the event rate is lower than expected on interim analysis of randomized trials, common strategies to adjust include increasing the sample size to compensate or to lengthen the study duration in the hope of capturing more events. Strategies can involve changing the inclusion or exclusion criteria, adding more study centers, or increasing the study duration (1). However, in general, investigators must be cognizant that altering the inclusion or exclusion criteria can result in the enrollment of participants with an event rate that is different (lower or higher) from the original inclusion cohort and different from participants in the prior studies used to inform study design. The inclusion criteria in PEXIVAS included patients who could possibly be at lower risk of the composite outcome, *i.e.*, patients with a higher eGFR (3,4). An interim analysis subsequently predicted a 5-year event rate that was lower than expected (34% versus 38%). In order to achieve the proposed event rate, a protocol amendment was introduced to increase recruitment from 500 to 700 participants. Substantial protocol amendments performed mid-trial can affect data analysis and interpretation. Transparency in the conduct and reporting of protocol amendments is recommended, including a review by an independent party such as the institutional review board. Although no prespecified process for dealing with protocol amendments was outlined in version 1.1 of the PEXIVAS protocol, the rationale behind the re-evaluated sample size calculation is clearly stated in subsequent protocol versions.

Fundamental to all sample size calculations is the decision by investigators on the likely size of the treatment effect (the  $\Delta$ ) that is attributable to the experimental intervention being studied. Usually, this is based on the minimal effect of the treatment that would be considered clinically relevant: “the minimally relevant treatment difference” (1,2). A downward adjustment is often made of this projected  $\Delta$  to build resilience against crossover of intervention arms and possible lags between treatment and effect. In addition, the sample size is often inflated in order to compensate for a lower-than-expected event rate due to competing and censoring events, while full compliance to the intervention is assumed (1,2). Estimating a more conservative  $\Delta$  will naturally lead to a larger sample size recommendation, which will help protect the design against not having sufficient sample size to detect smaller treatment effects attributable to the intervention under investigation. This  $\Delta$  metric, along with the likely expected event rate, is usually informed by the available literature, ideally from previous randomized clinical trials or high-quality observational studies in the same context.

The authors of PEXIVAS stated that they chose a large treatment effect size because PLEX is expensive and is invasive in nature (6). Rather than asking if there is any possible treatment benefit to PLEX, randomized trials such as PEXIVAS sometimes desire instead to ask if there is a treatment effect size ( $\Delta$ ) large enough to justify the invasiveness, risks of adverse events, and expense of PLEX and one that is worthwhile pursuing across the entire worldwide population of patients with acute presentations from ANCA vasculitis (generalizable). The original PEXIVAS protocol suggested that the sample size was calculated based on a 12% absolute risk reduction for PLEX on the primary composite outcome event of death or ESRD (number needed to treat= $1/\text{absolute}$

risk reduction]). It was powered based on a median time to ESRD or death of 6 years, enrolment over 5 years, and maximum follow-up of 7 years (6). To detect a hazard ratio of 0.64, investigators estimated that 490 patients were required to achieve 80% power, allowing for 10% loss to follow-up in both groups (6). The interim analysis conducted during PEXIVAS highlighted an event rate that was lower than expected and resulted in an inflation of sample size recruitment from 500 to 700 participants. Choosing a large threshold for  $\Delta$  due to the invasive nature of PLEX is very reasonable; however, it raises the possibility that the power of the trial was inadequate to detect a smaller treatment effect (a type II error)—a fact acknowledged by the authors (7).

The inability of PLEX to meet the predefined threshold for statistical significance does not rule out a treatment effect of a smaller size than the prespecified threshold. This is particularly true when the event rate is lower than expected (1). Common reasons for event rates being lower than expected include period effects and the enrollment of participants who are at lower risk of the outcome than originally expected. A period effect refers to the issue whereby outcome event rates change over the passage of time due to multifactorial reasons such as improvements in general and specific medical care for the disease over time. We created Figure 1, C1 and C2 to illustrate event rates aligning with given relative hazards to help readers grasp the effect of event rates on detecting effects of varying magnitude (8). Although direct *post hoc* assessments of achieved power of randomized trials are generally not very informative, it can be illustrative and educational to hypothesize how bigger and smaller  $\Delta$  treatment effects might be evaluated by various event rates (Figure 1). Although the result of PEXIVAS suggest confidence in the absence of a large overall treatment effect in a heterogeneous acute ANCA cohort, more moderate effects and effects in particular subgroups are still possible. This is particularly relevant when we consider subgroups of patients who may have derived benefit from PLEX, such as patients requiring dialysis at presentation, patients with active crescents on biopsy, and patients with pulmonary hemorrhage (7,9). Taken together, the results suggest that PEXIVAS may not have been adequately powered to detect small to moderate effects in this heterogeneous ANCA cohort.

In conclusion, the choice of an appropriate sample size is fundamental to the power of a study. PEXIVAS adjusted for censoring and competing risks in their initial sample size estimates by conducting a time-to-event analysis. However, a mid-trial analysis of PEXIVAS demonstrated an event rate that was lower than expected. This low event rate was potentially mediated by the inclusion criteria of participants at lower risk of the composite outcome and possibly by period effects among other possibilities. Despite the adjustment of the sample size estimate to capture the expected event rate, based on the observed event rate, the possibility of beneficial treatment effects in subgroups of patients presenting with ANCA vasculitis remains. PEXIVAS was a landmark study in nephrology, which contains a myriad of learning opportunities for researchers and clinicians in nephrology.

#### Disclosures

D.J. Sexton reports being on the advisory board for AstraZeneca and Takeda. The remaining author has nothing to disclose.

### Funding

D.J. Sexton is funded by Health Research Board (HRB) grant HRB-ARPP-P-2018-011. V. Sandys is funded by Enterprise Ireland Disruptive Technologies Innovation Fund (DTIF) grant DT 2019 0086.

### Author Contributions

D.J. Sexton was responsible for conceptualization, formal analysis, investigation, methodology, software, supervision, and validation; and both authors were responsible for visualization, wrote the original draft of the manuscript, and reviewed and edited the manuscript.

### References

1. Friedman LM, Furberg CD, DeMets DL: *Fundamentals of Clinical Trials*, New York, Springer, 2010
2. Peacock J, Peacock P: *Oxford Handbook of Medical Statistics*, 1st Ed., Oxford, Oxford University Press, 2010
3. Walsh M, Merkel PA, Peh CA, Szpirt WM, Puéchal X, Fujimoto S, Hawley CM, Khalidi N, Floßmann O, Wald R, Girard LP, Levin A, Gregorini G, Harper L, Clark WF, Pagnoux C, Specks U, Smyth L, Tesar V, Ito-Ihara T, de Zoysa JR, Szczeklik W, Flores-Suárez LF, Carette S, Guillevin L, Pusey CD, Casian AL, Brezina B, Mazzetti A, McAlear CA, Broadhurst E, Reidlinger D, Mehta S, Ives N, Jayne DRW; PEXIVAS Investigators: Plasma exchange and glucocorticoids in severe ANCA-associated vasculitis. *N Engl J Med* 382: 622–631, 2020 <https://doi.org/10.1056/NEJMoa1803537>
4. Jayne DR, Gaskin G, Rasmussen N, Abramowicz D, Ferrario F, Guillevin L, Mirapeix E, Savage CO, Sinico RA, Stegeman CA, Westman KW, van der Woude FJ, de Lind van Wijngaarden RA, Pusey CD; European Vasculitis Study Group: Randomized trial of plasma exchange or high-dosage methylprednisolone as adjunctive therapy for severe renal vasculitis. *J Am Soc Nephrol* 18: 2180–2188, 2007 <https://doi.org/10.1681/ASN.2007010090>
5. Riffenburgh RH: *Statistics in Medicine*, 3rd Ed., London, Elsevier Academic Press, 2012
6. Walsh M, Merkel PA, Peh CA, Szpirt W, Guillevin L, Pusey CD, De Zoysa J, Ives N, Clark WF, Quillen K, Winters JL, Wheatley K, Jayne D; PEXIVAS Investigators: Plasma exchange and glucocorticoid dosing in the treatment of anti-neutrophil cytoplasm antibody associated vasculitis (PEXIVAS): Protocol for a randomized controlled trial. *Trials* 14: 73, 2013 <https://doi.org/10.1186/1745-6215-14-73>
7. Walsh M, Merkel PA, Jayne DRW: Plasma exchange and glucocorticoids in severe ANCA-associated vasculitis. Reply. *N Engl J Med* 382: 2169, 2020 <https://doi.org/10.1056/NEJMoa1803537>
8. UCSF Clinical and Translational Science Institute: Sample Size Calculators for Designing Clinical Research. Available at: <http://www.sample-size.net/sample-size-survival-analysis/>. Accessed March 15, 2022
9. Nezam D, Porcher R, Grolleau F, Morel P, Titeca-Beauport D, Faguer S, Karras A, Solignac J, Jourde-Chiche N, Maurier F, Sakhi H, El Karoui K, Mesbah R, Carron PL, Audard V, Ducloux D, Paule R, Augusto J-F, Aniot J, Tiple A, Rafat C, Beaudreuil S, Puéchal X, Gobert P, Massy Z, Hanrotel C, Bally S, Martis N, Durel C-A, Desbuissons G, Godmer P, Hummel A, Perrin F, Néel A, De Moreuil C, Goulenok T, Guerrot D, Grange S, Foucher A, Deroux A, Cordonnier C, Guilbeau-Frugier C, Modesto-Segonds A, Nochy D, Daniel L, Moktefi A, Rabant M, Guillevin L, Régent A, Terrier B: Kidney histopathology can predict kidney function in ANCA-associated vasculitides with acute kidney injury treated with plasma exchanges. *J Am Soc Nephrol* 33: 628–637, 2022 35074934

**Received:** August 2, 2021 **Accepted:** May 20, 2022