How to Cite this article: Vicki Sandys and Donal Sexton, Learning methodological lessons from exemplar studies in Nephrology: PEXIVAS and sample size calculation, Kidney360, Publish Ahead of Print, 2022, 10.34067/KID.0005052021

Article Type: Clinical Research Methods

Learning methodological lessons from exemplar studies in Nephrology: PEXIVAS and sample size calculation

DOI: 10.34067/KID.0005052021

Vicki Sandys and Donal Sexton

Key Points:

Abstract:

Disclosures: D. Sexton reports the following: Advisory Board: AstraZeneca, Takeda. V. Sandys has nothing to disclose.


Author Contributions: Vicki Sandys: Visualization; Writing - original draft; Writing - review and editing; Donal Sexton: Conceptualization; Formal analysis; Investigation; Methodology; Software; Supervision; Validation; Visualization; Writing - original draft; Writing - review and editing

Data Sharing Statement:

Clinical Trials Registration:

Registration Number:

Registration Date:

The information on this cover page is based on the most recent submission data from the authors. It may vary from the final published article. Any fields remaining blank are not applicable for this manuscript.
Learning methodological lessons from exemplar studies in Nephrology:

PEXIVAS and sample size calculation

Vicki Sandys¹ and Donal J Sexton²,³

1. Royal College of Surgeons in Ireland,
   Dublin,
   Ireland
2. St James's Hospital,
   Dublin 8,
   Ireland.
3. School of Medicine,
   Trinity College Dublin,
   Dublin,
   Ireland

Corresponding author:

Professor Donal J Sexton,
Renal Unit, St James's Hospital,
Dublin 8,
Ireland.

Email dosextontcd.ie
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.\textsuperscript{3,4} Sample size calculation depends upon several factors, including the pre-specified 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.3,4 These estimates can be inferred prior to 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.3 In this short education piece, we examine the impact of sample size
estimates using PEXIVAS as an exemplar study.

Plasma exchange and glucocorticoid dosing in the treatment of anti-neutrophil
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.1 Prior to the publication of PEXIVAS, the use of PLEX
and higher dose corticosteroids in ANCA vasculitis was widespread in induction
protocols.1,2 The results of PEXIVAS demonstrated no benefit of PLEX on the
primary composite outcome of death and ESRD (hazard ratio 0.86, 95% CI: 0.65-
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.3-5

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. 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 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.

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. However, one of the interesting issues which can arise if 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 study duration. 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. 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 pre-specified 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 which would be considered clinically relevant, “the minimally relevant treatment difference”.3,4 A downward adjustment is often made of this projected delta to build resilience against cross-over 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. 3,4 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. 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/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. 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. 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.

The inability of PLEX to meet the predefined threshold for statistical significance does not out-rule a treatment effect of a smaller size than the pre-
specified threshold. This is particularly true when the event rate is lower than expected. 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 (Panels C1&C2) to illustrate event rates aligning with given relative hazards to help readers grasp the impact of event rates on detecting effects of varying magnitude. 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). While the result of PEXIVAS suggest confidence in the absence of a large overall treatment effect in a heterogenous acute ANCA cohort, more moderate effects are still possible, as well as effects in particular subgroups. This is particularly relevant when we consider subgroups of patients that may have derived benefit from PLEX, such as patients requiring dialysis at presentation, patients with active crescents on biopsy, and patients with pulmonary hemorrhage. Taken together, the results suggest that PEXVIAS 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. Sexton reports the following: Advisory Board: AstraZeneca, Takeda. V. Sandys has nothing to disclose.

Funding
DJS is funded by Health Research Board (HRB) grant HRB-ARPP-P-2018-011
VS is funded by Enterprise Ireland Disruptive Technologies Innovation Fund (DTIF) grant DT 2019 0086

Author Contributions
Vicki Sandys: Visualization; Writing - original draft; Writing - review and editing.
Donal Sexton: Conceptualization; Formal analysis; Investigation; Methodology; Software; Supervision; Validation; Visualization; Writing - original draft; Writing - review and editing.
References


Figure 1.

*Panels A1-A2 & B1-B2* refers to a hypothetical situation whereby projected event rates for a new study with a time to event primary outcome (*Panels B1 & B2*) are based on event rates from a historical study (*Panels A1 & A2*) with participants from the same study population (for example IgA Nephropathy). One can appreciate how an event rate lower than expected (3/10 in *Panel B2* versus 6/10 in *Panel 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). $t_{1-t_{15}}$ represents time from 1st follow up visit to the last in a hypothetical study.

*Panels C1-C2*. In PEXIVAS the event rate corresponds to 80% power to detect a relative hazard ratio of 0.679 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. $\beta$ = probability of Type II error.