CLINICAL INVESTIGATION

Novel adaptive designs: aligning drug development and patient incentives

"What does work, is a vision of showing how patient interests and company portfolio interests are aligned..."

Keywords: group sequential design • novel adaptive design • sample size reestimation

John W Loewy
DataForeThought, 52 Hillcrest Pky, Winchester, MA 1890, USA
jwloewy@comcast.net

In the November 2014 issue of the TIRS journal Stephen Spielberg [1] wrote an editorial motivating the future direction of the journal. He remarked: '*We are making progress in drug development, with faster, smarter and more effective ways of evaluating the efficacy and safety of new products, all of which is beginning to transform into better regulatory practices with alignment in high-quality, validated science.*' Also, '*We need to share together what works and what does not, all in the interest of patients in need and for all who strive to develop important new products.*'

The aim of this piece is to begin a discussion to focus on sharing what works and what we could be doing to work more effectively when the objective is to consider the implementation of the Novel Adaptive Design (NAD) clinical trial as a pivotal Phase III trial in a regulatory setting.

The NAD is continuing to gain favor as an alternative strategy for implementation of a pivotal trial in a regulatory setting. By 'novel,' we mean those adaptive designs that provide for an unblinded interim analysis in order to implement a predefined adaptive design decision. Since the issuance of the draft FDA adaptive design guidance [2] in 2010 we have come to understand that methods that provide for an unblinded interim analysis to estimate the treatment effect distinguishes the less-well understood Novel Adaptive Design (NAD) from the well-understood Adaptive Design [2,3] (AD).

To clarify further, 'less-well understood' designs tend to entail using the unblinded treatment effect to potentially modify some aspect of the design while a trial continues. We can distinguish this from a conventional Group Sequential Design (GSD), which is characterized as 'well understood,' where the actions are built into the study design and typically only involve whether the study terminates or continues unchanged.

NAD is being advocated in R&D programs as tools for optimizing clinical trial portfolio management [4]. The potential for increasing the probability of success, accelerated timing for product approval and reduced trial costs represent the optimistic advocacy on behalf of such designs. Kenneth Getz writes: '*Adaptive trial designs hold promise in optimizing study design. Early study terminations due to futility and sample size reestimation could save up to a hundred million dollars (USD) in direct and indirect costs annually per pharmaceutical company depending on clinical trial scope, when the trial is actually terminated, and on the sponsor's overall implementation of this adaptive approach across the development portfolio* [5].'

If NAD for pivotal clinical trials hold the promise that is being advertised why are we not seeing more of them being introduced as the implementation method of choice?

Functional silos or a cross-functional team of committed researchers

In general, it has been the experienced statisticians representing industry, academia and the regulatory authorities, along with a group of clinicians specializing in trial design, who have spearheaded the principles of NAD. Increasingly, new books and publications [4,6–7] continue to advance the role of NAD in clinical development strategy.

FUTURE SCIENCE part of fsg

Others have been more cautious. The cornerstone of the debate can be described along the following lines:

- NAD clinical trials have many moving parts that need to be made clearer to senior management teams representing sponsors who will fund such trials in a regulatory environment that is not always predictable;
- Adaptations of sample size, dropping treatment arms and other aspects of study designs that impose physical adaptive changes to the ongoing trial are increasingly more noticeable when compared with the GSD and thus continue to heighten concerns about potential biases that can impact the study integrity: especially, if the study sponsor has access to unblinded interim analyses which are the basis for the adaptations being made;
- Such physical changes, even though they are made based on a pre-specified adaptive decision rule, are keeping otherwise very responsible groups of researchers held in abeyance to making in-roads on the NAD acceptance trajectory.

No company sponsor, regardless of what could be a sense of regulatory buy-in, wants to have a pivotal clinical trial design that could be viewed as 'less-well understood,' implemented and then second-guessed.

What works & the way forward

On November 11, 2014 in Helsinki, the World Medical Association (WMA) celebrated the 50th anniversary of the Declaration of Helsinki. After 50 years, the Declaration continues as a living document. The commitment of respecting the human dignity of all patients participating in experimental research remains at the central core of its mission and serves as a reminder that designing even a traditional randomized controlled trial has ethical issues that continue to be debated.

> "...the greatest value of the Novel Adaptive Design is not that the trial design will adapt but that it can adapt."

Caution for NAD implementation begins because cross-functional groups sometimes have great difficulty discussing what matters when attempting to plan pivotal trials using NAD. Words and phrases like 'faster,' 'smarter,' 'more effective ways of evaluating,' become platitudes for concepts that speak only in general sound bites lacking direction or purpose. Often, such jargon only becomes associated with the capacity of an adaptive trial to mitigate the financial risks of drug development, rather than the inherent dignity of patients for whom NAD are being strategized.

NAD trial thinking introduces new territory to explore. Stakeholders in cross-functional groups from all sides of this 'elephant' can no longer remain in their functional silos simply to repeat the tasks for which training has made them well prepared to be successful in the past. However, it is much easier to retreat into a functional silo if the goal is to mitigate the financial risks of drug development instead of what matters to the patients participating in the trials that we as clinical researchers would have them enter.

What does work, is a vision of showing how patient interests and company portfolio interests are aligned, and how an applicable NAD can provide for treating fewer patients on an experimental therapy that can potentially shorten the time of drug development for a product that will benefit patients. This is especially true for patient indications suffering with life-threatening diseases for which there is an unmet need for treatment alternatives.

An example of planning to determine whether to introduce an NAD or to choose the well-understood GSD

The provision for implementation of an unblinded sample size re-estimation (SSR) is one of the most common NAD submitted to the regulatory authorities. In practice, this adaptation has a multitude of potential impacts on all trial stakeholders based on the pre-specified decision rules:

- Increase the originally planned sample size (decreasing the size is not encouraged);
- Clinical equipoise reconsideration by investigators after a trial has executed a study design adaptation;
- Trial logistics to implement proper firewalls so that sensitive information is controlled for the purpose of minimizing any potential study bias;
- The false positive rate or statistical control of type 1 error.

SSR is typically presented under an NAD in order to increase the sample size to save a study from potentially being under powered. This framework can be misguided. The GSD may also be more advantageous if the goal is simply to preserve statistical power. Instead, discussions about NAD implementation are more likely to prosper in a cross functional setting if minimizing the number of patients to be treated on a less effective treatment arm is the shared incentive. However, it is crucial during trial planning, that all members of cross-functional teams, not just statisticians, understand how the proposed NAD translates into the desired

result. The cross-functional chemistry for weighting the risks and benefits for NAD implementation take on a new awareness when the statistical considerations for NAD implementation are made plain to all.

This can be done through a comparison of the interplay occurring between a GSD and an NAD SSR. The GSD is already the pillar of preplanning that is well-understood. All of the pre-specifications for such designs are able to be deduced from commercially available software or software packages found though open-source code [8]. Some examples of software packages to explore are:

- Software packages commercially available:
  - EAST™ 6.3 by Cytel Inc;
  - ADDPLAN™ v6.1 by Aptive Solution Company;
  - PASS 13: Hintze, J. (2014). PASS 13. NCSS, LLC. Kaysville, Utah, USA [9].
- R programming language [10]:
  - Module within the R programming language, Package denoted as gsDesign;
  - Module within the R programming language, Package denoted as RCTdesign.

Using oncology as an example, new indications for experimental treatments for patients who have experienced and failed first- and second-line regimens of prior treatments for metastatic disease can be randomized to phase III studies in which the treatment control arm is assumed to have a median time to overall survival on the order of 6 months for some indications and on the order of 12 months for others. Because of the unmet need in these patient populations, randomized placebo-controlled trials in which an experimental treatment is to be compared with a best supportive care control treatment arm plus a placebo can sometimes constitute an ethical undertaking. Even so, it would be desirable to treat as few patients as possible on the control arm if the new drug experimental arm showed overwhelming promising results during a preplanned interim analysis for which the overall power of the trial is 90% and the type 1 error is defined at $\alpha = 0.025$ for a one-sided log-rank test.

To use the interplay mentioned previously, we are able to combine standard type of GSD's with α and β spending rules in a pivotal trial setting in combination with the adaptive statistical information design [11], to illustrate a side-by-side comparison of implementing an NAD SSR as part of the trial compared with a GSD without an NAD SSR.

To illustrate, suppose that the median time to death in the control group was estimated to be 6 months and it is projected that the new experimental treatment could extend this median time to 10 months. A GSD pivotal trial designed with 90% power could be constructed such that approximately 164 death events would be required to detect a hazard ratio (HR) equal to 0.60 (6 months/10 months) using a one-sided statistical test at $\alpha = 0.025$. The null hypothesis is that HR = 1. Without loss of generality, suppose that the GSD provided for a preplanned O'Brien-Fleming Lan-DeMets α spending function. A β spending function can also be included as a parameter for completeness but is ignored in this example. It can be shown that if the second interim analysis was undertaken at approximately 64% of the preplanned deaths (approx. 105) then the rejection boundary for efficacy of the trial would occur if the HR was approximately less than or equal to 0.60 which is the HR for which the trial is designed.

Typically in drug development, we can't be precisely sure of the median time to death of the control treatment arm. So cautiously, keeping everything the same the median time to death in the control group could also be estimated to be 6.5 months instead of 6 months. The revised GSD pivotal trial designed with 90% power would now require approximately 231 death events to detect HR = 0.65 (6.5 months/10 months). Similarly, it can be shown that if the second interim analysis was undertaken at approximately 64% of the preplanned deaths (approx. 148) then the rejection boundary for efficacy follows the pattern of the first design. Here, HR = 0.65 is the boundary value for rejection. Both designs reject the null hypothesis with a p-value of approximately $p = 0.005$ at 64% of the preplanned death events.

Wang *et al.* show that the implementation of an NAD SSR has the potential to be type 1 error penalty-free if the maximum sample size (in this case the number of death events) under SSR is at most 1.5 times the original sample size and the conditional power is at least 50%. In the example above the effect size of an interim analysis at 64% of deaths could be as low as approximately HR = 0.73 based on standard calculations in order to have greater than 50% conditional power to undertake performing an SSR without requiring a type 1 error adjustment. This implies that the GSD designed with a final analysis based on 164 deaths could be increased to as many as 246 deaths without a type 1 error penalty. This number is greater than the number of death events in the final analysis of the GSD designed to detect HR = 0.65. Therefore, NAD SSR makes it possible to aggressively design a trial in the hopes

of not having to adapt but with the ability to adapt under circumstances that can be discussed without deep simulations up front to begin discussions. To summarize the high points of the example above:

- In reality it is simply too difficult to pinpoint with precision the trial design that minimizes the number of patients treated on a less favorable treatment arm in a GSD that needs to be fixed in advance. An NAD could make a difference;
- Boundaries for rejection are presented on a treatment effect scale rather than a statistical scale so that all stakeholders are better able to quickly understand the impact of the statistical considerations;
- Discussions about alternative clinical trial tradeoffs become increasingly intuitive when only 64% of the events are necessary to stop the trial for the pre-specified protocol treatment effect for which the trial is powered. This is because the p-value is approximately $p = 0.005$ when using the conservative α spending function at this level of information;
- If the median time to death is 6 months, then if the study is designed for an HR = 0.6, this implies that the median will be extended to roughly 10 months in the experimental treatment arm. A trigger of 105 deaths conveys the sample size of deaths required to be approximately 64% of the total number of death events for which the boundary for stopping the trial for efficacy can be discussed;
- If the trial were designed with an HR = 0.65, the total number of death events increases from 164 to 231. The 64% interim analysis increases from 105 to 148 deaths;
- Given the opportunity of performing an SSR at 64% of the events for a trial designed with an HR = 0.60 instead of an HR = 0.65, it could mean stopping a trial triggered for an interim analysis based on 105 events instead of 148 events if the treatment is effective;
- With the ability to perform the correction of an SSR during an interim analysis, both the patient perspective and the portfolio analysis perspective could be aligned without a statistical penalty if the conditional power of the interim analysis remained above 50% and the maximum sample size adjustment for events was on the order of 1.5 of the original trial design [11];
- SSR provides a mechanism for not having to overcompensate for a lack of knowledge up front by being able to design the study with an HR = 0.60 instead of 0.65. Further, not stopping the trial for efficacy does not mean that an SSR would occur at the interim analysis. But just staying the course using the study design based on the HR = 0.60 could mean the difference of reducing the need to randomize an additional 100 patients to obtain an extra 40 or so deaths before the trial is concluded;
- All of the above depends on the accrual assumptions and the time-to-event assumptions but all of the concepts above can be described and discussed in a planning phase that compares whether it is worthwhile to implement an NAD beginning with stakeholder discussions that are able to frame minimizing the number of patients required before doing any advanced simulation up front;
- Advanced simulation is a step that can be undertaken in concert with specific details of a project for which such work will add value to the final planning for implementation.

NAD SSR is simply an extension of the GSD, but the potential for real designs that will reduce the number of patients receiving ineffective treatment is undeniable in the right circumstances.

Discussion

Expanding the momentum for industry wide adoption of NAD will continue to increase through examples that can articulate optimizing clinical trial implementation strategy combined with the interests of patients being asked to participate in the trials designed by sponsors.

Cross-functional groups vested in clinical trial strategy need to be able to communicate the talking points that matter not only pertaining to the probability of success as defined in a decision analysis for optimal pharmaceutical research and development programs, but also, as the points pertain to the patient who signs the informed consent agreeing to participate in the randomized controlled clinical trial.

In cross-functional groups, statisticians have an opportunity to emphasize statistical effect sizes in terms of scales that are easily understood by others. It is the effect sizes that matter the most when trying to understand the meaningfulness of a treatment effect for which the trial is being designed. This must be translated into endpoint concepts that other members of the cross-functional groups will grasp in order to work on operational aspects of the implementation of an NAD that will matter.

In the example above the greatest value of the NAD is not that the trial design will adapt but that it can adapt. The toolkit for NAD SSR plans for an adaptation as a safeguard to being overambitious when

establishing the sample size for the trial instead of planning from a risk-averse perspective.

On page 468 of the book, 'The Emperor of all Maladies: A Biography of Cancer' [12] the author writes: 'Gleevec had turned out to be so effective that doctors could no longer justify treating GIST patients with a placebo pill. Germaine started on the drug in August 2001. A month later, her tumors began to recede at an astonishing rate. Her energy returned; her nausea vanished. She was resurrected from the dead.'

In the above, we are reminded of the reasons to remain steadfast in our resolve to look for ways to minimize the number of patients receiving ineffective or inferior treatments in a clinical trial setting. To that end, NAD needs to be probed further by all clinical trial stakeholders for the potential that is possible on behalf of the brave patients participating on clinical trials and for those not participating on a clinical trial, but who are waiting for the next Gleevec to be approved.

Acknowledgements

J Loewy would like to thank P Gallo, for helpful comments and discussions while writing this manuscript.

Financial & competing interests disclosure

J Loewy is an independent statistical consultant having affiliations with BioBridges, LLC and qPharmetra, LLC. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1 Spielberg S. "Whence a Journal…" *Ther. Innov. Regul. Sci.* 48(6) 669–670 (2014).

2 US Food and Drug Administration. Draft guidance for industry on adaptive design clinical trials for drugs and biologics. Silver Spring, MD (2010). www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation

3 Wang SJ. A Commentary on the US FDA adaptive design draft guidance and EMA reflection paper from a regulatory perspective and regulatory experiences. In: *Practical Considerations for Adaptive Trial Design and Implementation.* He W, Pinheiro J, Kuznetsova OM (Eds). Springer, NY, USA (2014).

4 Antonijevic Z. *Optimization of Pharmaceutical R&D Programs and Portfolios.* Antonijevic Z (Ed.). Springer, Switzerland (2015).

5 Getz K. Improving protocol design feasibility to drive drug development economics and performance. *Int. J. Environ. Res. Public Health* 11, 5069–5080 (2014).

6 Hung HMJ, Wang SJ, & Yang P. Some challenges with statistical inference in adaptive designs. *J. Biopharm. Stat.* 24(5), 1059–1072 (2014).

7 He W, Pinheiro J, Kuznetsova OM. *Practical Considerations for Adaptive Trial Design and Implementation.* He W, Pinheiro J, Kuznetsova OM (Eds). Springer, NY, USA (2014).

8 Tymofyeyev Y. A review of available software and capabilities for adaptive designs. In: *Practical Considerations for Adaptive Trial Design and Implementation.* He W, Pinheiro J, Kuznetsova OM (Eds). Springer, NY, USA (2014).

9 NCSS Statistical Software. www.ncss.com

10 R programming language. www.cran.r-projects.org

11 Wang SJ, Brannath W, Bruckner M, Hung HMJ, Koch A. Unblinded adaptive statistical information design based on clinical endpoint or biomarker. *Stat. Biopharm. Res.* 5(4), 293–310 (2013).

12 Mukherjee S. *Emperor of All Maladies: A Biography of Cancer.* Mukherjee S (Ed.). Scribner, NY, USA (2010).