



SONAR

Do a New Design and Statistically Significant Results Translate to Reliability?

Michael Walsh

CJASN 15: 889–891, 2020. doi: <https://doi.org/10.2215/CJN.08540719>

At the World Congress of Nephrology in 2019, two trials testing medications to reduce the risk of progressing diabetic kidney disease were presented. Almost unheard of in our field, both trials reported “positive” headline results, the Atrasentan and Renal Events in Patients with Type 2 Diabetes and CKD Trial (SONAR) and Canagliflozin and renal outcomes in type 2 diabetes and nephropathy (CREDENCE) trial (1,2). SONAR tested whether atrasentan, an selective endothelin A receptor antagonist, reduced the incidence of the primary outcome of a doubling of serum creatinine, ESKD, or kidney death in patients with diabetic kidney disease that experienced at least a 30% reduction in albuminuria when treated with open-label atrasentan during a run-in period (2). The trial was stopped early with only about half the planned sample size (and outcome events) because of slow recruitment, but ultimately the trial reported a significant lower risk of the primary (kidney) outcome in the atrasentan group (hazard ratio, 0.65; 95% confidence interval, 0.49 to 0.88; $P=0.005$). As with any trial, SONAR raises many questions, but in this case there are questions about the novel trial design and the credibility of the observed treatment effect. These questions merit further thought both because it is important to consider how atrasentan may play a role in the treatment of diabetic kidney disease and because innovative designs in randomized, controlled trials are likely to be more commonly used.

In any randomized, controlled trial, eligible participants should be those likely to benefit from the experimental treatment. Testing the effects of treatments in those unlikely to benefit exposes those participants to unnecessary risks and reduces the probability of identifying benefits. Unfortunately, medicine is still relatively crude in how “likelihood to benefit” is determined. Mostly, a range of pathophysiologies are assigned a diagnosis and, on the basis of this imprecise diagnosis, we consider patients eligible for a treatment (after excluding patients for whom the treatment would be likely dangerous). For example, in the treatment of myocardial infarctions with streptokinase, patients with coronary syndromes owing to vasospasm or dissection were unlikely to benefit from streptokinase, a drug for thromboses. However, it was difficult to differentiate these diagnoses from myocardial

infarctions caused by plaque rupture and thrombosis within the time frame the drug was thought to be useful. Some patients with a pathology unlikely to benefit may therefore have been included but the trials still found streptokinase was, on average, effective presumably because such patients were a small proportion of the total (3).

Ideally, investigators pair pathophysiology with a treatment’s known mechanism(s) of action. This highlighted in modern oncology where malignancies have specific characteristics that make them susceptible to particular drugs. For example, the discovery that concomitant therapy with a mAb against HER-2 improved the survival of patients with HER-2–positive breast cancer (4). Now malignancies are being increasingly screened for markers that provide clues to their susceptibilities to particularly therapies. However, it is important to recognize that these approaches still have limitations and biomarker guided therapies, no matter how logical the treatment approach seemed, still do not always improve on less precise strategies.

SONAR was designed on the basis of the observation that risk of progressive kidney disease is correlated to albuminuria. Therefore, a drug that reduces albuminuria in some patients (a biologic response) but not in others may be more likely to exert a clinically important benefit in those with a biologic response. Enrolling only those with a biologic response makes sense as it would spare some the side effects and burden of a treatment unlikely to benefit them, while improving statistical power for the trial. There are caveats to the usefulness this design. Most importantly, the biologic response is most likely to predict a clinical treatment response if the biologic response lies on the causal pathway. History tells us that our assumptions that associations between biologic responses and clinical response, no matter how consistent and strong, are causal, are frequently incorrect (5). Second, if a biologic response is rare, the usefulness of the therapy to the population as a whole is likely limited. If a biologic response is almost universal, the extra step to determine a response is wasteful. If the biologic response is very delayed, participants are exposed to treatment-related harms that are not taken into account in the randomized part of the trial. Finally, if there is substantial measurement error in the biologic response, it will simply reduce the number of eligible

Population Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, Canada
Department of Medicine and Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Canada

Correspondence:
Dr. Michael Walsh, St. Joseph’s Hospital, 50 Charlton Avenue, Hamilton, ON L8N 4A6, Canada. Email: lastwalsh1975@gmail.com

patients rather than add statistical power. Regardless of these issues, the design limits the generalizability of the results to patients with a biologic response. In SONAR's case, the results are only generalizable to treated patients whose albuminuria falls $\geq 30\%$ after starting atrasentan (which is not the same as those in which atrasentan caused the reduction, because some changes are because of random variation).

None of these caveats fundamentally reduce the validity of the findings of the SONAR trial. The most important part of this strategy is that a biologic response is assessed before randomization. If the trial differentially treated or followed participants on the basis of response after randomization, the results would be at high risk of bias. So, the design is valid, but limits the application of the results to patients with $\geq 30\%$ drop in albuminuria after starting atrasentan.

Importantly, we need to consider the effect of excluding participants that did not experience a reduction in albuminuria. The SONAR investigators, wisely, studied these patients and in their presentation but not in the article, the 1020 nonresponders did not have dissimilar results to the responders. This suggests that including only participants with an albuminuric response may not adequately identify those most likely to benefit.

Once we accept that the design of SONAR is valid, we need to decide if the observed difference in outcomes was truly the effects of the drug. This is conventionally done by ensuring the difference is unlikely attributable to the play of chance (*i.e.*, statistically significant), that the risk of bias was reasonably low, and that the totality of evidence is sufficiently consistent (6).

From the perspective of ensuring the play of chance was unlikely, SONAR meets conventional criteria. The *P* value for the hypothesis test is low (not just <0.05 but $<0.005!$) and the 95% confidence interval excludes the null result. However, the fallibility of these metrics, and particularly treating them as binary is well documented (7). Simulation studies on the basis of cardiovascular outcome trials suggest treatment effects may be severely misestimated unless they use at least 600 outcome events (8). In the SONAR trial, there were 184 primary outcome events. Another way of thinking about this is with the Fragility Index, the number of events added to the intervention group that are required for the *P* value to slip from <0.05 to ≥ 0.05 (9). In the case of SONAR, only one extra event in the atrasentan group would potentially change the statistical inference in SONAR. A difference of one event is not difficult to imagine—it could happen by a random occurrence of a low eGFR, such as one patient taking extra nonsteroidal anti-inflammatory drugs or having a diarrheal illness in the placebo group, or one death occurring in the atrasentan group after a patient starts dialysis but before 90 days elapse. It is also notable that although the statistical significance would be lost by a different outcome in a single patient, 369 patients did not complete the study, of which 43 were lost to follow-up, and their outcomes could have easily contributed that one event difference. Similarly, RCTs stopped early for benefit, not the case with SONAR, consistently overestimate treatment effects (10). Taken together, these suggest there is insufficient data to make firm conclusions about whether atrasentan truly benefits patients or not.

In terms of risk of bias, SONAR had all the design elements to minimize bias: allocation was concealed, the investigators, patients, care providers and outcome adjudicators were all blinded, and follow-up and outcome ascertainment were excellent and similar in both groups.

Finally, consider the consistency of the results of SONAR, internally and with other trials. Internally, the results are broadly similar across the various definitions of the kidney outcomes and across the subgroups. Externally, there is little to compare with in terms of drugs in this class as they did not progress to the same stage of development. However, if we consider other effective drugs for preventing progression of diabetic kidney disease, both angiotensin receptor blockers and SGLT-2 inhibitors may give us an idea about expected treatment effects. Angiotensin receptor blockade reduces ESKD and doubling of creatinine by about 21% (11). Canagliflozin reduced a similar outcome by about 30%. In reality, the SONAR investigators powered the trial to detect (and therefore expected) a 27% reduction and the need for 425 outcome events. Although not unimaginable, the finding of a 35% reduction may stretch the limits of plausibility.

On the whole, the SONAR investigators should be commended for designing a trial that should have improved how we test and ultimately prescribe a new, promising treatment. However, because of its early termination, whether atrasentan truly reduces the risk of progressive diabetic kidney disease remains in question, as is whether future trials should use similar design. The use of a response in albuminuria may ultimately have created a logistical barrier to trial recruitment but without adding much in the way of power, and may pose a knowledge translation issue for future results. Although we certainly need to put careful thought into how we move forward with randomized trials, decades-old wisdom may still be very relevant today, and questions around common conditions are most reliably answered with a large, simple trial.

Acknowledgments

Dr. Walsh is supported by a Mid-Career Research Award from McMaster University.

The content of this article does not reflect the views or opinions of the American Society of Nephrology (ASN) or CJASN. Responsibility for the information and views expressed therein lies entirely with the author(s).

Disclosures

Dr. Walsh has nothing to disclose.

Funding

None.

References

- Perkovic V, Jardine MJ, Neal B, Bompoint S, Heerspink HJL, Charytan DM, Edwards R, Agarwal R, Bakris G, Bull S, Cannon CP, Capuano G, Chu PL, de Zeeuw D, Greene T, Levin A, Pollock C, Wheeler DC, Yavin Y, Zhang H, Zinman B, Meininger G, Brenner BM, Mahaffey KW; CREDESCENCE Trial Investigators: Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med* 380: 2295–2306, 2019
- Heerspink HJL, Parving HH, Andress DL, Bakris G, Correa-Rotter R, Hou FF, Kitzman DW, Kohan D, Makino H, McMurray

- JJV, Melnick JZ, Miller MG, Pergola PE, Perkovic V, Tobe S, Yi T, Wigderson M, de Zeeuw D; SONAR Committees and Investigators: Atrasentan and renal events in patients with type 2 diabetes and chronic kidney disease (SONAR): A double-blind, randomised, placebo-controlled trial. *Lancet* 393: 1937–1947, 2019
3. Intravenous streptokinase in acute myocardial infarction. *N Engl J Med* 315: 1356–1357, 1986
 4. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, Swain SM, Pisansky TM, Fehrenbacher L, Kutteh LA, Vogel VG, Visscher DW, Yothers G, Jenkins RB, Brown AM, Dakhil SR, Mamounas EP, Lingle WL, Klein PM, Ingle JN, Wolmark N: Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 353: 1673–1684, 2005
 5. Baker SG, Kramer BS: A perfect correlate does not a surrogate make. *BMC Med Res Methodol* 3: 16, 2003
 6. Walsh M, Perkovic V, Manns B: Therapy (randomized trials). In: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*, edited by Guyatt G, Rennie D, Meade MO, Cook DJ, 3rd Ed., New York, NY, McGraw-Hill Education, 2015
 7. Wasserstein RL: *ASA Statement on Statistical Significance and P-Values*, Alexandria, VA Amer Statistical Assoc, 2016
 8. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, Guyatt G, Devereaux PJ, Thabane L: The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis—a simulation study. *PLoS One* 6: e25491, 2011
 9. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, Molnar AO, Dattani ND, Burke A, Guyatt G, Thabane L, Walter SD, Pogue J, Devereaux PJ: The statistical significance of randomized controlled trial results is frequently fragile: A case for a fragility Index. *J Clin Epidemiol* 67: 622–628, 2014
 10. Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, Heels-Ansdell D, Walter SD, Guyatt GH, Flynn DN, Elamin MB, Murad MH, Abu Elnour NO, Lampropoulos JF, Sood A, Mullan RJ, Erwin PJ, Bankhead CR, Perera R, Ruiz Culebro C, You JJ, Mulla SM, Kaur J, Nerenberg KA, Schünemann H, Cook DJ, Lutz K, Ribic CM, Vale N, Malaga G, Akl EA, Ferreira-Gonzalez I, Alonso-Coello P, Urrutia G, Kunz R, Bucher HC, Nordmann AJ, Raatz H, da Silva SA, Tuche F, Strahm B, Djulbegovic B, Adhikari NK, Mills EJ, Gwadrý-Sridhar F, Kirpalani H, Soares HP, Karanicolas PJ, Burns KE, Vandvik PO, Coto-Yglesias F, Chrispim PP, Ramsay T; STOPIT-2 Study Group: Stopping randomized trials early for benefit and estimation of treatment effects: Systematic review and meta-regression analysis. *JAMA* 303: 1180–1187, 2010
 11. Strippoli GF, Bonifati C, Craig M, Navaneethan SD, Craig JC: Angiotensin converting enzyme inhibitors and angiotensin II receptor antagonists for preventing the progression of diabetic kidney disease. *Cochrane Database Syst Rev* (4): CD006257, 2006

Published online ahead of print. Publication date available at www.cjasn.org.