Is the Mortality Benefit With Empagliflozin in Type 2 Diabetes Mellitus Too Good To Be True?

Sanjay Kaul, MD

Signaling a likely end to a long and elusive quest for cardiovascular outcome benefit associated with treatment intervention in type 2 diabetes mellitus, the results of the EMPA-REG OUTCOME trial [BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients] were received with a standing ovation at the European Association for the Study of Diabetes scientific meeting in Stockholm, Sweden, on September 17, 2015.[1] Witnessing the spontaneous applause, I had mixed emotions. Was it time to bring the trumpets out and rejoice that the "holy grail" had finally been achieved? Or, was it more appropriate to curb the enthusiasm and question the "historic milestone," given that the mortality benefit was unexpected and unprecedented?

Examples abound of instances where we have been led astray by implausibly large treatment effects that were not confirmed by subsequent trials. Perhaps the most compelling is the case of perioperative β-blockade with bisoprolol in high-risk vascular surgery.[2] The DECREASE 1 trial (Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echo) yielded a 91% risk reduction in cardiovascular death or myocardial infarction ($P<0.001$) in 112 patients. These results were widely disseminated and adopted by several practice guidelines, ultimately rising to the status of a performance measure. The positive results of this trial were never replicated. On the contrary, a large, randomized trial (POISE [Perioperative Ischemic Evaluation]) and a meta-analysis pointed to harm, necessitating a downgrading of recommendations a decade after the publication of the original trial results.[3] One systematic review concluded that most large treatment effect estimates should be considered with caution. The vast majority are either spurious findings or represent substantial overestimations, and large mortality benefits are almost entirely nonexistent.[4]

Thus, the key question that lingered in my mind despite the resounding applause was, "Should we simply dismiss these unexpected results to be 'too good to be true' and attribute them to a play of chance?" In answering this question, I wrestled with the following arguments.

First, both all-cause mortality and cardiovascular mortality were prespecified as secondary end points, although they were not included in the statistical hierarchical testing strategy, which included a stepwise evaluation of noninferiority followed by superiority of 3 and 4-point major adverse cardiovascular events (MACE). A purist might argue that because superiority of 4-point MACE was not met ($P=0.079$), the α error had already been spent, and therefore all subsequent analyses, including mortality, must be deemed exploratory, requiring confirmation in subsequent trials. Taken to a logical extreme, this is akin to saying that because Christopher Columbus had prespecified discovering a route to India, America must not exist. There is regulatory precedence of a successful claim of carvedilol reducing the combined incidence of morbidity and mortality in heart failure despite the fact that mortality was not prespecified as a primary or a secondary end point in the pivotal trials.[5]

Correspondence to: Sanjay Kaul, MD, Division of Cardiology, Cedars-Sinai Medical Center, 8635 W Third St, Ste 790W, Los Angeles, CA 90048. E-mail kaul@cshs.org

Key Words: clinical trial ■ diabetes mellitus ■ mortality

Second, the mortality benefit is large and clinically important: 2.6% absolute or 32% relative risk reduction in all-cause mortality and 2.2% absolute or 38% relative risk reduction in cardiovascular mortality; and statistically robust ($P<0.001$ for both). These results are consistent with the quantity of evidence necessary to support the US Food and Drug Administration's substantial evidence criterion of effectiveness based on a single trial that requires a highly persuasive statistical finding (ie, $P<0.001$).

Third, the mortality benefit is based on large number of events: 463 all-cause and 309 cardiovascular mortality events. Remarkably, <1% of patients had missing information on vital status.

Fourth, a consistent mortality benefit is seen with both doses: 30% and 33% relative risk reduction in all-cause mortality and 35% and 41% relative risk reduction in cardiovascular mortality with the 10- and 25-mg dose, respectively.

Finally, the *P* values for both all-cause and cardiovascular mortality are robust enough to preserve type 1 or false-positive error after adjustment for >100 comparisons. It is important to emphasize that both all-cause and cardiovascular mortality, but not 3-point MACE (the primary end point), results satisfy the key attributes of regulatory decision making: prespecification, replication, and preservation of type 1 error.

The shortcomings of using *P* values as a measure of evidence are well documented and continue to stir much controversy.[6] Some have argued that *P* values overestimate the strength of the evidence and offered the use of Bayes factor, which is a measure of how well the null and the alternative hypotheses predict the data.[6] The minimum Bayes factor and the corresponding strength of evidence for 3-point MACE and mortality results in EMPA-REG OUTCOME are shown in Table 1. The *P* value of 0.038 for 3-point MACE translates into a minimum Bayes factor of 0.131, which means that the evidence supports the null hypothesis approximately one eighth as strongly as it does the alternative. This reduces the null probability from 50% before the trial to 10% after the trial. This does not represent strong evidence against the null, thus requiring independent confirmation in a subsequent trial. For all-cause and cardiovascular mortality, the nominal *P* value of 0.0001 translates into Bayes factors of 0.0006 (1/1815) and 0.0004 (1/2358), which reduces the extremely skeptical prior null probability of 95% to <0.5% after the trial, indicating very strong evidence against the null.

A formal Bayesian analysis[7] of EMPA-REG OUTCOME shown in Table 2 provides useful insights. For all-cause mortality, the posterior hazard ratio (HR) shifts from 0.68 using the noninformative prior to 0.76 (95% confidence interval excluding an HR of 1), which is still a clinically important treatment effect. Similarly, for cardiovascular mortality and hospitalization for heart failure, the posterior HR shifts from 0.62 to 0.74 and from 0.65 to 0.80, respectively (95% confidence interval excludes an HR of 1). For the 3-point MACE, the HR shifts from 0.86 to 0.88 (95% confidence interval no longer excludes an HR of 1). One can also estimate the probability of a range of treatment effects. Thus, if one deems 15% mortality reduction as the minimum clinically important difference, then the probability of achieving this is 99% using the noninformative prior and 92% using the skeptical prior. Thus, by for-

Table 1. Evaluating Strength of Evidence of Cardiovascular Outcomes in EMPA-REG OUTCOME Using Bayes Factor

End Point	*P* Value (*z* Score)	Minimum Bayes Factor	Decrease in Probability of Null Hypothesis, %		Strength of Evidence	Effect Size, HR
			From	To No Less Than		
3-Point MACE	0.038 (2.02)	0.131	95	54	Moderate	0.86
			75	28		
			50	12		
All-cause mortality	0.0001 (3.94)	0.0006	95	0.49	Very strong	0.68
			75	0.16		
			50	0.06		
Cardiovascular mortality	0.0001 (3.87)	0.0004	95	0.38	Very Strong	0.62
			75	0.13		
			50	0.04		
Hospitalization for heart failure	0.0017 (2.93)	0.0137	95	11	Strong	0.65
			75	4		
			50	1		

Bayes' theorem: posterior odds=prior odds x evidence (Bayes factor). Bayes factor=prob (data/H0)/prob (data/H1) (likelihood ratio); H0=null hypothesis; H1=alternative hypothesis. Minimum Bayes factor=$\exp(-0.5z^2)$. Odds=probability/(1−probability). Probability=Odds/(1+Odds).

EMPA-REG Outcome indicates BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients; HR, hazard ratio; and MACE, major adverse cardiovascular event.

Table 2. Bayesian Analysis of Cardiovascular Outcomes in EMPA-REG OUTCOME

End Point	Prior	Evidence	Posterior	Probability of Benefit		
				*P*b≥0	*P*b≥10%	*P*b≥15%
3-Point MACE	Noninformative	0.86 (0.74–0.99)	0.86 (0.74–0.99)	0.980	0.767	0.506
	1.00 (0.75–1.33) (skeptical)	0.86 (0.74–0.99)	0.88 (0.70–1.01)	0.966	0.619	0.304
All-cause mortality	Noninformative	0.68 (0.57–0.82)	0.68 (0.57–0.82)	0.999	0.998	0.992
	1.00 (0.75–1.33) (skeptical)	0.68 (0.57–0.82)	0.76 (0.65–0.89)	0.999	0.982	0.916
Cardiovascular mortality	Noninformative	0.62 (0.49–0.77)	0.62 (0.49–0.77)	0.999	0.999	0.997
	1.00 (0.75–1.33) (skeptical)	0.62 (0.49–0.77)	0.74 (0.62–0.89)	0.999	0.982	0.925
Hospitalization for heart failure	Noninformative	0.65 (0.50–0.85)	0.65 (0.50–0.85)	0.999	0.989	0.971
	1.00 (0.75–1.33) (skeptical)	0.65 (0.50–0.85)	0.80 (0.65–0.97)	0.988	0.884	0.728

Bayesian analysis allows information from earlier trials, if available, (the prior) to be integrated with the current evidence (likelihood) to generate a posterior.[7] Two types of prior are used: (1) Noninformative or vague prior: all effect sizes are equally plausible (log OR=0, log SD=10). The choice of noninformative prior can be reasonably justified, reflecting the uncertainty associated with the possible benefit of empagliflozin therapy. In this case, the posterior is driven entirely by the evidence (as in the frequentist approach); (2) Skeptical prior: mean OR=1; 95% CI, 0.75–1.33 (probability of OR<0.75 is 2.5% and OR>1.33 is 2.5%; log OR=-0.001, log SD=0.146). Probability of at least 0, 10% and 15% reduction in outcome is shown.

MACE indicates major adverse cardiovascular event.

mally incorporating skepticism, the Bayesian approach helps moderate results that are "too good to be true."

Critics have argued that lack of a clear and biologically plausible mechanism underlying mortality benefit is a major limitation. This is a rather uncharitable criticism because outcome trials are not designed to unravel the potential mechanisms of benefit. What we can say with reasonable confidence from the trial results so far is that mortality benefit is unlikely to be mediated by favorable but very modest effects on cardiometabolic factors such as blood pressure, body weight, or glycemic control, given the rapid onset of treatment effect (curves separate as early as 2–3 months), and it is unlikely to be mediated by an atherothrombotic effect, given the lack of effect on myocardial infarction and stroke. The observations that hospitalization for heart failure was reduced by 35% and that half of the cardiovascular mortality advantage was driven by reduction in worsening heart failure and sudden cardiac death[1] support a possible hemodynamic or antiarrhythmic effect. Future studies aimed at these targets should help yield mechanistic insights.

Thus, the totality of data suggests that the observed magnitude of mortality benefit in EMPA-REG OUTCOME is not likely to be spurious. Nonetheless, because the findings were unexpected and unprecedented and not linked to obvious mechanistic pathway, the results need to be replicated in future investigations. Only then can we be sure beyond any reasonable doubt that the mortality results are highly reliable and that it is time to take the trumpets out to herald the historic milestone.

DISCLOSURES

Dr Kaul has a consultant or advisory relationship with Boehringer-Ingelheim, sponsor of empagliflozin, and Eli Lilly, collaborator with Boehringer-Ingelheim for empagliflozin.

AFFILIATION

From Division of Cardiology, Cedars-Sinai Medical Center, Los Angeles, CA.

FOOTNOTES

Circulation is available at http://circ.ahajournals.org.

REFERENCES

1. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE; EMPA-REG OUTCOME Investigators. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med*. 2015;373:2117–2128. doi: 10.1056/NEJMoa1504720.
2. Poldermans D, Boersma E, Bax JJ, Thomson IR, van de Ven LL, Blankensteijn JD, Baars HF, Yo TI, Trocino G, Vigna C, Roelandt JR, van Urk H. The effect of bisoprolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography Study Group. *N Engl J Med*. 1999;341:1789-1794.
3. POISE Study Group, Devereaux PJ, Yang H, Yusuf S, Guyatt G, Leslie K, Villar JC, Xavier D, Chrolavicius S, Greenspan L, Pogue J, Pais P, Liu L, Xu S, Málaga G, Avezum A, Chan M, Montori VM, Jacka M, Choi P. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet*. 2008;371:1839-1847. doi: 10.1016/S0140-6736(08)60601-7.
4. Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. *JAMA*. 2012;308:1676–1684. doi: 10.1001/jama.2012.13444.
5. Fisher LD, Moyé LA. Carvedilol and the Food and Drug Administration approval process: an introduction. *Control Clin Trials*. 1999;20:1–15.
6. Goodman SN. Toward evidence-based medical statistics, 2: the Bayes factor. *Ann Intern Med*. 1999;130:1005–1013.
7. Kaul S. Are concerns about reliability in the trial to assess chelation therapy fair grounds for a hasty dismissal? An alternative perspective. *Circ Cardiovasc Qual Outcomes*. 2014;7:5–7. doi: 10.1161/CIRCOUTCOMES.113.000714.