A Comparison of Individuals with Diabetes and EMPA-REG Trial Participants: Exploring Aspects of External Validity

Alexander Chaitoff, MD, MPH[1] , Joshua D. Niforatos, MD, MTS[2], Jingyi Gong, MD[1], and Michael A. Fischer, MD, MS[3]

[1]Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, 75 Francis St, Boston, MA, USA; [2]Department of Emergency Medicine, The Johns Hopkins School of Medicine, Baltimore, MD, USA; [3]Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Harvard Medical School, Boston, USA.

BACKGROUND: There is increasing use of sodium glucose co-transporter 2 (SGLT2) inhibitors to treat diabetes. Since trials apply specific entry and exclusion criteria to ensure internal validity, comparisons of trial populations with nationally representative samples can inform the applicability of study findings to practice.

OBJECTIVE: To compare individuals with diabetes from a nationally representative sample to patients who underwent randomization in the EMPA-REG trial. A secondary aim was to characterize what proportion of individuals prescribed an SGLT2 inhibitor in a nationally representative sample would have been included in the EMPA-REG trial.

DESIGN: Retrospective cross-sectional study.

PARTICIPANTS: Adults with diabetes who took part in the National Health and Nutrition Examination Survey (NHANES) between 2011–2014 (primary analysis corresponding to EMPA-REG enrollment) and 2015–2018 (secondary analysis corresponding to contemporary sample).

MAIN MEASURES: The primary outcome was a comparison of demographic (age, sex, ethnicity, and pregnancy status), clinical (comorbidities and medication use), examination (weight, body mass index, and systolic and diastolic blood pressure), and laboratory (hgba1c, low- and high-density lipoprotein cholesterol, triglycerides, and estimated glomerular filtration rate) characteristics of NHANES respondents versus EMPA-REG trial participants. The secondary outcome was the proportion of NHANES respondents who had been prescribed an SGLT2 inhibitor that would have met inclusion criteria for the EMPA-REG trial.

KEY RESULTS: There were 655 and 48 respondents, representing a weighted sample of 21,849,775 and 1,062,573 individuals, included in the primary and secondary analyses, respectively. Overall, 7.6% (95% CI 4.8–10.6%) of 2011–2014 NHANES respondents would have met all EMPA-REG trial inclusion criteria. NHANES respondents and EMPA-REG participants differed across demographic, clinical, examination, and laboratory domains. Of NHANES respondents from 2015 to 2018 who were prescribed an SGLT2 inhibitor, 10.6% (95% CI <1–24.7%) would have met all inclusion criteria for the EMPA-REG trial.

CONCLUSIONS: The EMPA-REG population differed from a nationally representative sample, which could affect generalizability.

INTRODUCTION

Applying randomized clinical trial (RCT) results to patients in routine clinical practice presents challenges. RCTs typically focus on selecting inclusion and exclusion criteria to maximize internal validity, potentially at the expense of external validity.[1] This design feature, compounded by often inadequate reporting of trial participant characteristics[2] and time pressures on providers,[3, 4] can make it difficult for clinicians to go deeper than general guidelines when necessary to identify which patients should receive a treatment in real-world settings. Ultimately, the multiple inclusion and exclusion criteria required to select such populations are typically presented in trial appendices or methods papers that are unlikely to be seen by the busy clinician.

Considering some of these limitations, there is increasing interest in using large real-world data sources to conduct observational research that can augment results from RCTs and guide evidence-based practice.[5] It has been suggested that studies using real-world data could be reported in addition to RCTs to provide the generalizability that is often sacrificed in RCTs that seek to maximize internal validity.[6, 7] However, a recent study on the feasibility of using real-world data to replicate RCTs suggests that currently available data can only support this approach for a fraction of trials.[8] Furthermore, there are well-known limitations to observational studies, including the challenges of conducting research with data collected for other purposes and the difficulty of reproducing results across different settings or data sources.[9–13] Accordingly, RCTs remain the "gold standard" for clinical evidence even though discordances between the population in which an

intervention is studied and the population in which it is used can have real consequences.[14] Specifically, average risks and benefits in the real-world may differ from those observed in trials, which could at the very least affect shared decision-making conversations.[15]

Recently, several major clinical trials have shown the effectiveness of sodium glucose co-transporter 2 (SGLT2) inhibitors for treating diabetes and cardiovascular comorbidities, resulting in the incorporation of this drug class into recent American Diabetes Association Guidelines.[16] However, further study is necessary to examine how closely the trial populations match the real-world patients who may be offered SGLT2 inhibitor therapy, and so this study had one primary and one secondary exploratory aim regarding just one of these trials studying SGLT2 inhibitors. Primarily, we sought to compare the characteristics of individuals living with diabetes to the baseline characteristics of patients who underwent randomization in the EMPA-REG trial, and secondarily, we hoped to characterize what proportion of individuals actually prescribed an SGLT2 inhibitor would have been included in the EMPA-REG trial.[17]

METHODS

Study Population

The sample consisted of adults aged 18 years and older who took part in the National Health and Nutrition Examination Survey (NHANES). NHANES is a cross-sectional survey conducted by the Centers for Disease Control and Prevention. NHANES uses a complex, multistage, clustered probability method to sample non-institutionalized US civilians to provide nationally representative estimates.[18]

To address the first aim of comparing patients who underwent randomization in the EMPA-REG trial with the general population with diabetes, we used responses from individuals sampled between 2011 and 2014 to correspond with the EMPA-REG study enrollment period. Respondents were eligible for inclusion if they had been selected to complete the interview, exam, and fasting laboratory sections of the survey. Respondents were included if they had diabetes, defined as having a glycated hemoglobin (hgba1c) $\geq 6.5\%$ or fasting glucose ≥ 126 mg/dL, and were not missing data. To address our second aim of determining the proportion of individuals that were prescribed SGLT2 inhibitors that would have met inclusion criteria for the EMPA-REG trial, we used responses from individuals sampled between 2015 and 2018, which are the most recent NHANES cycles available. Respondents were eligible for inclusion if they had completed the interview, exam, and laboratory sections of the survey. Respondents were included if they were prescribed a SGLT2 inhibitor and were not missing data.

Outcome Measures

Outcomes were defined based on demographic, clinical, laboratory, and medication variables. Demographic variables included age (years), sex (male vs female), ethnicity (Hispanic vs non-Hispanic), and pregnancy status (yes vs no). Clinical variables included heart disease (defined as self-reported history of angina, coronary artery disease, stroke, myocardial infarction, or heart failure), weight (kilograms), BMI (kg/m^2), and systolic and diastolic blood pressure (mmHg). Laboratory variables included hgba1c (%), low-density lipoprotein cholesterol (mg/dL), high-density lipoprotein cholesterol (mg/dL), triglycerides (mg/dL), and estimated glomerular filtration rate (GFR) as calculated by the MDRD equation (mL/min/1.73^2).[19] Medication variables included the number of individuals using prescription contraception, taking anti-hyperglycemic agents (including biguanides, sulfonylureas, dipeptidyl peptidase-4 (DPP-4) inhibitors, thiazolidinediones, SGLT2 inhibitors, and insulin), anti-hypertensive agents (including angiotensin-converting enzyme inhibitors/angiotensin II receptor blockers (ACEi/ARB), beta-blockers, diuretics, calcium-channel blockers, mineralocorticoid receptor antagonists, and renin inhibitors), and lipid-lowering therapies (including statins, fibrates, ezetimibe, and niacin). Medication usage was extracted from NHANES preferentially using Lexicomp drug class codes and with generic drug identification codes when no specific drug class existed. All data from NHANES were extracted using the *nhanesA* package in R while data from the EMPA-REG trial were extracted from Table S2 in the supplementary materials.[17]

The primary outcome, created based on these variables, was whether a respondent would have met inclusion or exclusion criteria for EMPA-REG. Criteria were extracted from the EMPA-REG trial protocol. EMPA-REG inclusion criteria included hgba1c $\geq 7.0\%$ and $\leq 10\%$ for patients on background therapy or hgbA1c $\geq 7.0\%$ and $\leq 9.0\%$ for drug-naïve patients, BMI ≤ 45 kg/m^2, and history of heart disease while exclusion criteria included pregnant women or women of child bearing age not on birth control, GFR ≤ 30 mL/min/1.73 (via the MDRD equation per trial protocol)[2], and substantially abnormal laboratory test (operationalized for this study as aspartate aminotransferase, alanine aminotransferase, or alkaline phosphatase greater than three times the upper limit of normal).[20] The secondary outcome was whether individuals actually prescribed an SGLT2 inhibitor from 2015 to 2018 would have met these same criteria.

Statistical Analysis

Descriptive statistics were used to describe sample means and the number of individuals who would have met inclusion criteria for the trial. One-sample *T*-tests were used to compare continuous NHANES sample means with continuous population means reported in the EMPA-REG trial. Chi-square goodness of fit tests were used to compare categorical

NHANES sample means with categorical population means reported in the EMPA-REG trial. Finally, a subgroup analysis, using one-sample T-tests and chi-square goodness of fit tests as described, was also conducted among only those who reported having heart disease. All analyses were conducted in R version 4.0.2 (https://cran.r-project.org/) or SAS (Cary, NC: SAS Institute Inc.) using appropriate survey weighting with the *survey* package in R or the SURVEYFREQ procedure, respectively. A two-sided alpha level of 0.05 was used to assess for statistical significance.

RESULTS

For the analysis of whether patients with diabetes would have been eligible for EMPA-REG, 655 respondents were included in the analysis, representing a weighted sample of 21,849,775 individuals with diabetes from 2011 to 2014. For the analysis of patients receiving SGLT-2 inhibitors in real-world settings, 48 respondents were included in the analysis, representing a weighted sample of 1,062,573 individuals who received an SGLT2 inhibitor from 2015 to 2018.

Compared with individuals in the EMPA-REG trial, NHANES respondents were younger (59.2 vs 63.1 years, $p<0.001$), were more likely to be female (45.6% vs 28.5%, $p<0.001$), weighed more (94.1 kg vs 86.3 kg, $p<0.001$), had a higher BMI (33.2 kg/m^2 vs 30.6 kg/m^2, $p<0.001$), and had lower systolic (131.1 mmHg vs 135.4 mmHg, $p<0.001$) and diastolic (70.7 mmHg vs 76.6 mmHg, $p<0.001$) blood pressure (Table 1). Among laboratory measures, NHANES respondents had lower hgba1c (7.55% vs 8.07%, $p<0.001$), higher LDL (102.8 mg/dL vs 85.7 mg/dL, $p<0.001$) and HDL (47.4 mg/dL vs 44.4 mg/dL, $p<0.001$), but lower triglycerides

(152.7 mg/dL vs 170.6 mg/dL, $p=0.002$), and higher GFR (84.4 mL/min/1.73^2 vs 74.0 mL/min/1.73^2, $p<0.001$). Finally, NHANES respondents were less likely to be on every class of anti-hyperglycemic and anti-hypertensive medication evaluated as well as less likely to be on statins (but not less likely to be on other lipid-lowering medications, Table 2).

When the analysis was limited only to NHANES respondents with cardiovascular disease, patterns of differences between the NHANES populations and the EMPA-REG populations remained similar for sex, weight, BMI, systolic and diastolic blood pressure, hgba1c, and LDL. However, the NHANES respondents with cardiovascular disease were less likely to be Hispanic (7.9% vs 18.0%, $p=0.001$) and did not have significantly different levels of HDL, triglycerides, or GFRs compared with EMPA-REG trial participants. Lastly, similar to patterns observed when comparing the total NHANES population with the EMPA-REG population, rates of medication use by NHANES respondents with cardiovascular disease were generally significantly lower than rates of medication usage by the EMPA-REG trial participants.

Overall, 7.6% (95% CI 4.8–10.6%) of 2011–2014 NHANES participants with diabetes would have met EMPA-REG trial inclusion criteria. Most notably, only about one-third of individuals would have met the hgba1c and drug cutoff criteria (36.6%, 95% CI 32.1–41.1%) and only a quarter would have met the high cardiovascular risk criteria (25.1%, 95% CI 20.1–30.2%; Table 3).

In the secondary analysis of individuals who participated in NHANES from 2015 to 2018 and were prescribed an SGLT2 inhibitor, 10.6% (95% CI <1–24.7%) would have met all inclusion criteria to be included in the EMPA-REG trial. Most NHANES respondents prescribed an SGLT2 inhibitor failed

Table 1 Comparison of NHANES and EMPA-REG Demographic, Exam, and Laboratory Characteristics

		NHANES		EMPA-REG
		Value (95% CI)		**Value (*N* or SD)**
		Total sample N=655)	*Sample with cardiovascular risk (N=153)*	*Total sample (N=7020)*
Age, years		59.2 (57.8, 60.5)***	66.2 (63.7, 68.7)*	63.1 (8.7)
Sex, %				
	Female	45.6 (40.2, 51.0)**	36.9 (25.2, 48.5)**	28.5 (2004)
Ethnicity, %				
	Hispanic	15.6 (10.7, 20.4)	7.9 (3.4, 12.5)**	18.0 (1265)
Weight, kg		94.1 (91.0, 97.2)***	95.2 (88.4, 102.1)*	86.3 (19.0)
BMI, kg/m^2		33.2 (32.4, 34.0)***	33.2 (31.4, 34.9)**	30.6 (5.2)
Systolic blood pressure, mmHg		131.1 (128.8, 133.4)***	130.9 (127.7, 134.1)*	135.4 (17.0)
Diastolic blood pressure, mmHg		70.7 (69.2, 72.2)***	65.4 (61.6, 69.3)**	76.6 (9.8)
Glycated hemoglobin, %		7.55 (7.35, 7.74)***	7.65 (7.06, 8.24)**	8.07 (0.85)
Low-density lipoprotein cholesterol, mg/dL		102.8 (99.0, 106.5)***	97.4 (90.9, 103.9)**	85.7 (35.8)
High-density lipoprotein cholesterol, mg/dL		47.4 (45.9, 48.8)***	45.6 (43.0, 48.3)	44.4 (11.7)
Triglycerides, mg/dL		152.7 (142.4, 163.0)**	163.0 (141.4, 184.6)	170.6 (126.9)
Estimated glomerular filtration rate, mL/min/1.73^2		84.4 (81.4, 87.4)**	70.8 (64.9, 76.7)	74.0 (21.4)

*$p<0.05$, **$p<0.01$, ***$p<0.001$

Table 2. Comparison of NHANES and EMPA-REG Medication Characteristics

	NHANES		EMPA-REG
	% (95% CI)		% (N)
	Total sample (N=655)	Sample with cardiovascular risk (N=153)	(N=7020)
Anti-hyperglycemic			
Metformin	44.3 (40.0, 48.6)***	41.2 (29.4, 54.4)**	74.0 (5193)
Sulfonylureas	25.6 (18.0, 32.6)***	35.7 (23.8, 47.5)	42.8 (3006)
DPP-4 Inhibitors[a]	8.2 (5.1, 11.4)	7.9 (2.2, 13.6)	11.3 (796)
Thiazolidinedione	3.6 (1.7, 5.5)	3.9 (<0.1, 8.0)	4.3 (299)
Insulin	23.5 (18.3, 28.7)***	33.6 (22.6, 44.7)**	48.2 (3387)
Anti-hypertensive			
ACEi/ARB[b]	52.2 (48.9, 55.7)***	68.1 (60.6, 75.6)***	80.7 (5666)
Beta-blocker	27.5 (22.0, 33.1)***	51.2 (39.6, 62.8)**	64.8 (4554)
Diuretics	27.8 (23.8, 31.8)***	37.7 (30.4, 45.0)	43.2 (3035)
Calcium-channel blocker	15.8 (11.1, 20.5)***	26.6 (17.1, 36.1)	33.0 (2317)
Mineralocorticoid receptor antagonists	3.5 (1.5, 5.4)**	4.8 (0.6, 8.9)	6.3 (441)
Renin inhibitors	<0.01 (<0.1, 0.5)	NA	0.6 (46)
Lipid-lowering			
Statins	45.3 (41.3, 49.2)***	64.2 (52.1, 76.4)**	77.0 (5403)
Fibrates	7.1 (3.5, 10.6)	15.3 (5.2, 25.4)	9.0 (630)
Ezetimibe	2.2 (0.6, 3.9)	2.7 (<0.1, 6.6)	3.8 (270)
Niacin	2.1 (0.2, 4.1)	4.0 (<0.1, 9.6)	1.8 (126)

*p<0.05, **p<0.01, ***p<0.001. [a]Dipeptidyl peptidase 4 inhibitors. [b]Angiotensin-converting enzyme inhibitors/angiotensin II receptor blocker

to meet the hgba1c and drug cutoff criteria (49.2%, 95% CI 24.1–74.4%) or the high risk of heart disease criteria (32.7%, 14.7–50.6%, Table 4).

Table 3 Description of Proportion of NHANES Population Meeting Major EMPA-REG Inclusion Criteria

Criteria	Meets criteria	
	N	% (95% CI)
Meets hemoglobin A1c and number of medication cutoff	230	37.4 (33.1, 41.8)
Meets BMI ≤45 kg/m^2	612	91.4 (87.9, 94.8)
Meets high cardiovascular risk	153	25.1 (20.1, 30.2)
Meets liver function tests less than than 3× the upper limit of normal	641	98.1 (96.7, 99.4)
Meets GFR >30 mL/min	642	98.0 (96.6, 99.5)
Meets not being pregnant and being on birth control is of child bearing age	610	93.1 (90.5, 95.7)

Table 4 Description of Proportion of NHANES Population Who Received an SGLT2 Inhibitor and Met Major EMPA-REG Inclusion Criteria

	Meets criteria	
	N	% (95% CI)
Meets hemoglobin a1c and number of medication cutoff	26	53.7 (28.9, 78.4)
Meets BMI ≤45 kg/m^2	44	96.8 (92.9, 99.9)
Meets high cardiovascular risk	12	32.3 (14.9, 49.8)
Meets liver function tests not more than than 3× the upper limit of normal	48	NA
Meets GFR >30 mL/min	48	NA
Meets not being pregnant or be at risk of becoming pregnant	46	96.2 (90.7, 99.9)

DISCUSSION

This study compared individuals with diabetes from the general population with participants from the EMPA-REG trial. Overall, individuals with diabetes in the NHANES sample differed substantially from EMPA-REG trial participants across demographic, clinical, laboratory, and medication domains. When only individuals with heart disease were considered, the NHANES sample was more similar to EMPA-REG trial population, but there were still significant differences across all domains. A minority of individuals with diabetes in the 2011–2014 NHANES sample, and a minority of individuals in a more contemporary NHANES sample who were prescribed an SGLT2 inhibitor would have met eligibility for the EMPA-REG trial.

We examined several components of external validity that clinicians might consider during their routine practice. First, they may consider how well the patient characteristics of trial populations, such as age, sex, and race, match their own clinical populations. Second, they may consider the stage of disease or extent of disease control in the trial population as compared with those parameters in their patients. Third, they may consider whether the patterns of care and treatment in a trial protocol are similar to the way that they practice; for example, study participants may have more frequent follow-up testing or medication titration than would be expected in typical practice. Assessments of the external validity of clinical trials often focus on the first component: how closely the demographic characteristics of patients enrolled in trials reflect those of the general population. Proposed solutions to mismatches between trial and general populations typically include increased recruitment of under-represented groups of patients,[21, 22] which represents an important priority for

researchers. However, the second and third components of external validity have been less widely studied and may require different mitigation approaches. In addition to recruiting a more diverse population, addressing these other challenges may require additional detail in the reporting of trial results that goes beyond subgroup analyses to include descriptions of how the stage of disease and patterns of care for trial participants compare with local, state, or national trends, which might assist both individual clinicians with decisions as well as regional and national organizations with guideline-making. Going further, trialist could consider recruitment plans with the aim of more closely matching trial populations to clinical populations on more than just demographic variables.

While prior work has identified criteria for clinicians and researchers to consider when evaluating external validity,[23, 24] many trials do not present enough information for clinicians to make these assessments. For example, one study that examined nearly 200 Norwegian drug trials affecting general practice found that a majority did not report one or more important variables necessary for assessing external validity.[25] Even when enough information is available, evaluations of the external validity of trials have found that inclusion and exclusion criteria often limit trial generalizability. A review of cardiology, oncology, and mental health studies that sought to compare trial populations with real-world populations concluded that over 70% of trials were not representative of real-world populations.[26] Our findings are consistent with this previous literature on trial generalizability.

Recent evaluations of the generalizability of clinical trials of novel anti-diabetic medications have concluded that no trial perfectly matches real-world populations.[27–30] These studies have reported the proportion of individuals from the general population who would have been included in trials of anti-diabetes medications based largely on demographic characteristics and a limited set of laboratory measurements but report only limited comparisons of other aspects of the trial versus real-world populations, and few describe how characteristics of the trial populations compare with characteristics of individuals in the general population receiving these agents. Our study builds on this previous work in two ways: first, by providing a more comprehensive framework for comparing a nationally representative population with a trial population, and, second, by reporting how many individuals receiving an SGLT2 inhibitor would have been included in one of the landmark trials.

The direction of the differences between NHANES respondents and trial participants was not consistent; for example, the NHANES populations was younger, had lower blood pressures, and had higher GFRs compared with the EMPA-REG trial populations, but NHANES respondents also had higher average BMI. The finding that NHANES respondents were less likely to be on multiple classes of medications might suggest that the population is healthier than trial participants though we cannot exclude the possibility that NHANES respondents were receiving less aggressive diabetes care than trial participants especially considering EMPA-REG

investigators were encouraged to aggressively treat cardiovascular risk factors, like hypertension and hyperlipidemia, during the study period. Either of these explanations could be problematic for the generalization of the trial results, particularly if SGLT2 inhibitors interact, either positively or negatively, with other drug classes. Even when only individuals with high cardiovascular risk were included, which would more closely approximate the EMPA-REG population that was enriched for such individuals, many of these differences remained. It is plausible that these differences could influence the efficacy of SGLT2 inhibitors; for example, the EMPA-REG trial reported that SGLT2 inhibitors tended to have less effect in individuals with BMI over 30. These findings mirror similar analyses from Europe that have suggested that effect sizes seen in the data from trials of SGLT2 inhibitors may only apply to a small subset of individuals with diabetes.[31]

Clinical trials cannot be perfectly representative of the entire population; it is more efficient for them to select participants for whom a given therapy is most likely to have benefit.[32] Our finding that a minority of NHANES respondents who received an SGLT2 inhibitor would have met trial inclusion criteria for EMPA-REG highlights another potential concern with the generalizability of these trials. While the confidence intervals for our estimates were wide, these preliminary findings point to an area where ongoing surveillance in the form of observational studies could be useful to both quantify effect sizes in real-world (vs trial) populations and possibly help redirect prescribing to more appropriate populations.

This study has several limitations. First, because of coding differences for variables in NHANES and the EMPA-REG trial, some characteristics, such as race, could not be compared. Moreover, this study could not differentiate NHANES respondents diagnosed with type 1 versus type 2 diabetes, but because type 2 diabetes in over 17 times more prevalent, misclassifications were unlikely to significantly affect results.[33] Relatedly, it is possible that some individuals did not meet the hgba1c cutoff for either the primary or secondary analysis because of improved control on anti-diabetic regimens, specifically SGLT2 inhibitors, as it relates to the latter analysis. Additionally, cardiovascular risk in NHANES is assessed via self-report rather than the more detailed medical information available to trial recruiters, so some NHANES respondents may or may not have been classified as high cardiovascular risk according to trial protocols, though it is unclear if the average primary care clinician makes decisions with information closer to that contained in NHANES questions or the EMPA-REG trial data. Furthermore, the sample size of NHANES respondents receiving an SGLT2 inhibitor from 2015 to 2018 was small, limiting our ability to make comparisons. While it is possible that individuals can be prescribed SGLT2 inhibitors for multiple indications, it is less likely that that was the case in this analysis as our study period predates the trials of SGLT2 inhibitors for use in kidney disease and heart failure.[34, 35] Future studies should more fully characterize the demographic, clinical, laboratory, and medication differences between those prescribed SGLT2 inhibitors and those who were included in

these more recent trials. Finally, this study did not assess all inclusion and exclusion criteria from the EMPA-REG protocol, including some subjective considerations, such as whether a prospective trial participant would be able to make follow-up visits and might therefore overestimate the percent of individuals meeting inclusion criteria.

This study builds on the current literature by exploring in more detail the differences between patients included in the EMPA-REG trial and the general US population with diabetes. Our findings demonstrate how the trial population differed significantly from a nationally representative sample in multiple domains and that few individuals with diabetes or who receive SGLT2 inhibitors would have met EMPA-REG trial eligibility criteria, which highlights the more general difficultly in applying the literature to clinical decision-making. Clinical trialists might address this issue by selecting inclusion and exclusion criteria with the expressed purpose of closely matching the population in which the drug will be used and providing clearer comparisons of how the trial population compared with the general population in which the drug might be marketed. These results suggest that observational research comparing trial populations with relevant real-world populations across multiple classes of variables could provide useful data to support clinicians attempting to apply the results from a trial to their own patients. Finally, clinicians should be mindful of the possible limitations of trial generalizability when making prescribing decisions for individual patients.

Corresponding Author: Alexander Chaitoff, MD, MPH; Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, 75 Francis St, Boston, MA 02120, USA (e-mail: achaitoff@bwh. harvard.edu).

REFERENCES

1. Flather M, Delahunty N, Collinson J. Generalizing results of randomized trials to clinical practice: Reliability and cautions. *Clinical Trials.* 2006;3(6):508-12. https://doi.org/10.1177/1740774506073464

2. Malmivaara A. Generalizability of findings from randomized controlled trials is limited in the leading general medical journals. *J Clin Epidemiol.* Published online 2019. https://doi.org/10.1016/j.jclinepi.2018.11.014

3. Konrad TR, Link CL, Shackelton RJ, et al. It's about time: physicians' perceptions of time constraints in primary care medical practice in three national healthcare systems. *Med Care.* 2010;48(2):95.

4. Tsiga E, Panagopoulou E, Sevdalis N, Montgomery A, Benos A. The influence of time pressure on adherence to guidelines in primary care: an experimental study. *BMJ Open.* 2013;3(4):e002700.

5. Patorno E, Schneeweiss S, Gopalakrishnan C, Martin D, Franklin JM. Using Real-World Data to Predict Findings of an Ongoing Phase IV Cardiovascular Outcome Trial: Cardiovascular Safety of Linagliptin Versus Glimepiride. *Diabetes Care.* 2019;42(12):2204-2210.

6. Blonde L, Khunti K, Harris SB, Meizinger C, Skolnik NS. Interpretation and impact of real-world clinical data for the practicing clinician. *Adv Ther.* 2018;35(11):1763-1774.

7. Bennette CS, Nussbaum NC, Curtis MD, Meropol NJ. Using real-world cohorts to assess the generalizability and relevance of randomized clinical trials (RCTs). J Clin Oncol 2019;37(15_suppl):6540-6540. https://doi.org/10.1200/JCO.2019.37.15_suppl.6540

8. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence. *JAMA Netw Open.* 2019;2(10):e1912869-e1912869.

9. Smith NC. Replication studies: A neglected aspect of psychological research. *Am Psychol.* Published online 1970. https://doi.org/10.1037/h0029774

10. Moonesinghe R, Khoury MJ, Janssens ACJW. Most published research findings are false—but a little replication goes a long way. *PLoS Med.* 2007;4(2):e28.

11. Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet.* Published online 2001. https://doi.org/10.1038/ng749

12. Grimes DA, Schulz KF. False alarms and pseudo-epidemics: The limitations of observational epidemiology. *Obstet Gynecol.* Published online 2012. https://doi.org/10.1097/AOG.0b013e31826af61a

13. Ioannidis JPA, Chen J, Kodell R, Haug C, Hoey J. Why Most Published Research Findings Are False. *PLoS Med.* 2005;2(8):e124. https://doi.org/10.1371/journal.pmed.0020124

14. Collins R, Bowman L, Landray M, Peto R. The Magic of Randomization versus the Myth of Real-World Evidence. *N Engl J Med.* 2020;382(7):674-678. https://doi.org/10.1056/NEJMsb1901642

15. Van de Ven N, Gilovich T, Zeelenberg M. Delay, doubt, and decision: How delaying a choice reduces the appeal of (descriptively) normative options. *Psychol Sci.* 2010;21(4):568-573.

16. Association AD. Summary of Revisions: Standards of Medical Care in Diabetes—2020. *Diabetes Care.* 2020;43(Supplement 1):S4-S6.

17. Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med.* Published online 2015. https://doi.org/10.1056/NEJMoa1504720

18. Centers for Disease Control and Prevention, National Center for Health Statistics. The National Health and Nutrition Examination Survey. https://www.cdc.gov/nchs/nhanes/index.htm. Accessed 15 Oct 2021.

19. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Ann Intern Med.* 1999;130(6):461-470.

20. Lala V, Goyal A, Bansal P, Minter D. A. Liver function tests. *StatPearls.* StatPearls Publishing 2020. Retrieved October 15, 2021 from https://www.ncbi.nlm.nih.gov/books/NBK482489/.

21. Loree JM, Anand S, Dasari A, et al. Disparity of race reporting and representation in clinical trials leading to cancer drug approvals from 2008 to 2018. *Jama Oncol.* 2019;5(10):e191870-e191870.

22. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA.* 2004;291(22):2720-2726.

23. Dekkers OM, Elm E von, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. *Int J Epidemiol.* 2010;39(1):89-94.

24. Rothwell PM. External validity of randomised controlled trials:"to whom do the results of this trial apply?" *Lancet.* 2005;365(9453):82-93.

25. Brænd AM, Straand J, Klovning A. Clinical drug trials in general practice: how well are external validity issues reported? *BMC Fam Pract.* 2017;18(1):113.

26. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials.* 2015;16(1):495.

27. Nicolucci A, Candido R, Cucinotta D, et al. Generalizability of cardiovascular safety trials on SGLT2 inhibitors to the real world: implications for clinical practice. *Adv Ther.* 2019;36(10):2895-2909.

28. Wittbrodt ET, Eudicone JM, Bell KF, Enhoffer DM, Latham K, Green JB. Eligibility varies among the 4 sodium-glucose cotransporter-2 inhibitor cardiovascular outcomes trials: implications for the general type 2 diabetes US population. *Am J Manag Care.* 2018;24(8 Suppl):S138-45.

29. Castellana M, Procino F, Sardone R, Trimboli P, Giannelli G. Generalizability of sodium-glucose co-transporter-2 inhibitors cardiovascular outcome trials to the type 2 diabetes population: a systematic review and meta-analysis. *Cardiovasc Diabetol.* 2020;19(1):1-10.

30. Boye KS, Riddle MC, Gerstein HC, et al. Generalizability of glucagon-like peptide-1 receptor agonist cardiovascular outcome trials to the overall type 2 diabetes population in the United States. *Diabetes, Obes Metab.* 2019;21(6):1299-1304.

31. McGovern A, Feher M, Munro N, de Lusignan S. Sodium-glucose cotransporter 2 (SGLT2) inhibitor: comparing trial data and real-world use. *Diabetes Ther.* 2017;8(2):365-376.

32. Rothman KJ, Gallacher JEJ, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol.* 2013;42(4):1012-1014.

33. Bullard KM, Cowie CC, Lessem SE, et al. Prevalence of diagnosed diabetes in adults by diabetes type—United States, 2016. *Morb Mortal Wkly Rep.* 2018;67(12):359.

34. Heerspink HJL, Stefánsson B V, Correa-Rotter R, et al. Dapagliflozin in patients with chronic kidney disease. *N Engl J Med.* 2020;383(15):1436-1446.

35. McMurray JJ V, Solomon SD, Inzucchi SE, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med.* 2019;381(21):1995-2008.