Real-world cardiovascular effects of liraglutide: transportability analysis of the LEADER trial

Kevin Josey, PhD[1,2], Wenhui Liu, MS[1], Theodore Warsavage, MS[1], Morten Medici, PhD[3], Kajsa Kvist, PhD[3], Catherine G. Derington, PharmD, MS[4,5], Jane E.B. Reusch, MD[1,6], Debashis Ghosh, PhD[2], Sridharan Raghavan, MD, PhD[1,7]

1 US Department of Veterans Affairs Eastern Colorado Health Care System, Aurora, CO.
2 Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO.
3 Novo Nordisk A/S, Copenhagen, Denmark.
4 Division of Cardiology, Department of Medicine, University of Colorado School of Medicine, Aurora, CO.
5 Adult & Child Center for Outcomes Research & Delivery Science, University of Colorado Anschutz Medical Campus, Aurora, CO.
6 Division of Diabetes, Endocrinology, and Metabolism, Department of Medicine, University of Colorado School of Medicine, Aurora, CO.
7 Division of General Internal Medicine, Department of Medicine, University of Colorado School of Medicine, Aurora, CO.

Corresponding authors:
Sridharan Raghavan, MD, PhD
Rocky Mountain Regional VA Medical Center
1700 North Wheeling Street
Aurora, CO 80045
Sridharan.raghavan@va.gov

Kevin Josey, PhD
Colorado School of Public Health
13001 East 17th Place
Aurora, CO 80045
Kevin.josey@cuanschutz.edu

Abstract: 197 words
Word count: 3994
Figures: 4
Tables: 1

Abstract

Appropriate use of recently approved type 2 diabetes treatments depends on external validity of landmark clinical trials (RCTs) in real-world populations that may differ from trial participants. This study transported effect estimates from the Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results (LEADER) trial, a placebo-controlled RCT of liraglutide on cardiovascular outcomes, onto nested real-world cohorts within the Veterans Affairs (VA) healthcare system. Risk differences (RD) in survival outcomes, approximated using pseudo-observations of individual survival probabilities, were estimated with augmented inverse probability weighting after balancing baseline characteristics between RCT and target samples using approximate balancing weights. Transported effects of liraglutide compared to placebo on major adverse cardiovascular events (MACE) and all-cause mortality in veterans ("VA-weighted LEADER") consistently overlapped the treatment effects observed in LEADER: MACE RD at 3 years of 2.0% [95% CI 0.8, 3.2] in VA-weighted LEADER versus 1.6% [0.3, 2.9] in LEADER; all-cause mortality RD at 3 years of 1.5% [0.6, 2.4] in VA-weighted LEADER versus 0.9% [-0.09, 1.9] in LEADER. The benefits of liraglutide observed in LEADER generalized to veterans with diabetes -- real-world evidence that can guide diabetes treatment decisions and formulary policies for a high-risk population underrepresented in RCTs.

Introduction

Randomized clinical trials (RCTs) provide gold-standard evidence for evaluating medication efficacy but often have uncertain external validity, especially when RCT participants are not representative of potential real-world users of a medication.[1] This uncertain external validity can complicate treatment decision-making at the point of care and may underlie delays in uptake of new, effective therapies.[1,2] Furthermore, RCTs often motivate regulatory decisions, treatment guidelines, and medication access policies, particularly for new therapies, resulting in regulatory approval, guidelines, and policies that impact diverse real-world populations based on data from non-representative RCT samples.[3-5] The disconnect between individuals sampled into RCTs and real-world patient populations is particularly acute in type 2 diabetes, a highly prevalent condition affecting a heterogeneous population with a substantial expansion in treatment options over the last decade. For example, RCTs of glucagon-like peptide-1 receptor agonists (GLP-1RAs), potent agents for the treatment of type 2 diabetes and obesity[6,7] with demonstrated cardiovascular benefits,[6] have consistently under-enrolled relevant constituencies of potential medication users, leading to gaps in our understanding about the effectiveness of these drugs in many populations.[8-13] Moreover, many newer diabetes medications with cardiovascular benefits, including GLP-1RA, are expensive, making their use and accessibility sensitive to treatment coverage policies.

While the most direct method for assessing medication effects in specific target populations is to conduct an RCT within each population, this strategy is impractical, and often infeasible, due to high costs and extensive time requirements. Observational studies

using real-world data, collected after a medication is in routine use, offer an appealing alternative to RCTs but can be limited by confounding and sampling bias. To address these limitations, transportability analyses have emerged as an alternative, middle-ground solution.[14,15] Methods for transporting effect estimates integrate RCT and real-world data to estimate what the treatment effect would have been if individuals from the real-world target population had been enrolled in the trial.[16] These methods preserve the benefits of randomization while utilizing data on potential medication users without requiring real-world uptake. Thus, transportability analyses can contribute evidence of efficacy for new medications in representative populations and inform formulary policies for healthcare payors servicing real-world potential users of a new medication.

The objective of this study was to assess external validity of the Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results (LEADER) study – the first RCT to demonstrate cardiovascular benefits of a GLP-1RA (liraglutide)[17] – in a real-world population under-represented in trials. We transported the results of the LEADER study[17] to US Department of Veterans Healthcare System (VA) users with type 2 diabetes, a real-world patient population comprising a small minority in foundational RCTs of GLP-1RAs[6,17-19] despite having higher cardiovascular disease risk than non-veteran US adults with diabetes.[20,21] Methodologically, we extend transportability methods to survival outcomes using pseudo-observations within a doubly-robust framework,[14,22,23] providing a template for transporting RCT results with time-to-event endpoints to real-world target populations. The study provides real-world evidence to guide VA formulary policies for GLP-1RA in a high-risk population in which direct RCT data are lacking.

Methods

Conceptual Framework for Transportability

Several methodological strategies are commonly employed for effect transportation: (1) Weighting Methods (Inverse Probability Weighting - IPW), which adjusts for differences between the trial and target (VA) populations by assigning weights based on the likelihood of trial participation given observed covariates, reweighting the RCT sample to resemble the VA cohort; (2) G-Computation, which involves modeling the outcome as a function of treatment and covariates to predict outcomes under different treatment scenarios in the target population; and (3) Doubly-Robust (DR) Methods (e.g., Augmented IPW; AIPW), which combines weighting and outcome modeling, achieving robustness by requiring only one of these two models (propensity for trial inclusion or outcome regression) to be correctly specified. We focus on the latter method as it combines the other two methods and enables the operationalization of machine learning to provide more robust, data-driven estimates without sacrificing parametric rates of convergence (e.g. the central-limit theorem holds for DR estimators despite using models that do not necessarily confer root-n consistency).

Pseudo-observations for Survival Probabilities

Pseudo-observations, as introduced by Per Kragh Andersen and Pohar Perme, provide a method for estimating survival probabilities at a given time using a leave-one-out jackknife approach.[22] This technique is particularly useful in regression modeling where

direct estimation of individual survival probabilities may be challenging. For participant $i = 1,2,\ldots,n$, and given the trial data where the samples have survival data, let Ŝ(t) be the Kaplan-Meier estimator of the survival function at time t. The individual pseudo-observations are defined as: $\tilde{S}_{\mathrm{i}}(t) = n \cdot \hat{S}(t) - (n-1) \cdot \hat{S}^{-i}(t)$, where $\hat{S}(t)$ is the Kaplan-Meier estimate of survival at time t using all data, and $\hat{S}^{-i}(t)$ is the Kaplan-Meier estimate of survival at time t computed after removing individual i from the dataset. By implementing this pre-processing step, we can treat the survival pseudo-observations as quasi-continuous outcomes so that we may apply standard causal inference methods without having to fit complex censoring-adjusted time-to-event models. We computed pseudo-observations for individual survival probabilities at 6-month intervals from 6 to 54 months.

Identification of the Transported Average Treatment Effect

Let $Z_i \in \{0,1\}$ indicate whether a sample belongs to the trial ($Z_i = 1$) or to the target ($Z_i = 0$), $A_i \in \{0,1\}$ be the indicator for treatment with liraglutide, and $S_i(t,1)$ and $S_i(t,0)$ represent the potential survival probability pseudo-outcomes under treatment ($A_i = 1$) and control ($A_i = 0$), respectively, for $i = 1,2,\ldots,n$. Define $\boldsymbol{X}_i = \left(1, X_{i1}, X_{i2}, \ldots, X_{ip}\right)^T$ as the set of covariates that we will need to balance and condition our outcome regression upon. The parameter of interest is the target population average treatment effect on survival:

$$\theta_0(t) = E[S_i(t,1) - S_i(t,0) | Z_i = 0].$$

To estimate this quantity with the observed data requires four assumptions:

(1) Stable Unit Treatment Value Assumption (SUTVA), consisting of no interference (the exposure assignment of one person does not influence the potential outcome of others) and consistency (the observed outcome corresponds with the factual potential outcome);

(2) Exposure Ignorability: $S_i(t,a) \perp A_i | \boldsymbol{X}_i, Z_i = 1$, meaning the potential outcomes are independent of exposure when we condition on the confounders, specifically in the trial sample;

(3) Sample Exchangeability: the average of the difference between potential outcomes is exchangeable across the trial and the target sample,

$$E[S_i(t,1) - S_i(t,0) | \boldsymbol{X}_i, Z_i = 0] = E[S_i(t,1) - S_i(t,0) | \boldsymbol{X}_i, Z_i = 1];$$

(4) Positivity: $0 < \Pr\{Z_i = 1 | \boldsymbol{X}_i\} < 1$.

Approximate Balancing Weights

In observational studies and causal inference, balancing weights are frequently used to reweight samples for unbiased estimation of treatment effects. Approximate balancing weights extend these methods by directly optimizing for covariate balance while maintaining statistical efficiency. The estimator we implement requires estimated weights $\hat{\gamma}_i, i = 1,2, \dots, n$, which are constructed to modify the marginal distribution of covariates $\boldsymbol{X}_i$, particularly the distribution of any effect modifiers, so that the weighted covariate means in the trial sample ($Z_i = 1$) match those of the target population represented by a target sample ($Z_i = 0$).

This can be formalized as an optimization problem that identifies the weights that balance the covariate moments[24] within a standardized range δ_j with

$$\min_{\boldsymbol{\gamma}\in\Re^n} \sum_{i=1}^{n} Z_i f(\gamma_i)$$

$$\text{subject to } \left|\sum_{i=1}^{n} \gamma_i Z_i X_{ij} - \sum_{i=1}^{n} (1-Z_i) X_{ij}\right| \leq \delta_j \text{ for all } j = 1,2,\ldots,p$$

and for a given Bregman distance $f(\gamma_i)$ (e.g. entropy).[16,25] If we select $\delta_j = c \cdot \sigma_j$ where σ_j is the standard deviation of X_{ij} in the target population, then we restrict the absolute standardized mean difference between the weighted trial covariates and the covariates in the target sample to fall within a constant range c of zero. We set $c = 0.05$ to achieve near-exact balance while preserving a larger effective sample size than exact balancing would permit. The effective sample size

$$ESS = \frac{(\sum_{i=1}^{n} Z_i \hat{\gamma}_i)^2}{\sum_{i=1}^{n} Z_i \hat{\gamma}_i{}^2}$$

quantifies the information available after weighting.

Doubly-Robust Estimation

We constructed DR estimators to estimate treatment effects specific to the target population by transporting average treatment effects from the RCT to the VA target population.[14,23] The DR augmented inverse probability weighted estimator combines both outcome modeling and weighting to correct for potential model misspecifications. The estimated expected treatment effect in the target population (Z = 0) at time t is $\hat{\theta}_0(t) = \frac{1}{n_0}\sum_{i=1}^{n} \hat{\psi}_0[A_i, \boldsymbol{X}_i, Z_i, S_i(t)]$, where

$$\hat{\psi}_0[A_i, \boldsymbol{X}_i, Z_i, S_i(t)] = Z_i \hat{\gamma}_i [S_i(t) - \hat{\mu}_t(A_i, \boldsymbol{X}_i)] + (1 - Z_i)[\hat{\mu}_t(A_i = 1, \boldsymbol{X}_i) - \hat{\mu}_t(A_i = 0, \boldsymbol{X}_i)]$$

We used Super Learner to fit the outcome model $\hat{\mu}_t(A_i, \boldsymbol{X}_i) = \hat{E}[S_i(t)|A_i, \boldsymbol{X}_i, Z_i = 1]$ with an ensemble of prediction algorithms incorporating generalized linear model regressions, elastic net regressions, multivariate adaptive spline regressions, and random forest regressions.[26] Note that $S_i(t)$ is not observed, so we use $\tilde{S}_i(t)$ instead. By combining the weighting method with an outcome model, this approach remains consistent even if one of the two components (weights or outcome model) is misspecified.

Application to LEADER and VA Data

Individual-participant data from the LEADER study was shared with the study team through a data use agreement with Novo Nordisk A/S. LEADER randomized 9,340 patients with type 2 diabetes to liraglutide (1.8 mg daily or maximum tolerated dose) or placebo, with median follow-up of 3.8 years. Target population data were derived from the VA Corporate Data Warehouse (CDW). We translated the major inclusion and exclusion criteria of the LEADER trial onto the VA data to create a LEADER-eligible VA cohort (Cohort A), requiring individuals to have a history of atherosclerotic cardiovascular disease (ASCVD), age ≥50 years, and HbA1c ≥7% at baseline. Unlike LEADER, we excluded individuals with eGFR <15 ml/min/1.73m^2 as liraglutide prescribing patterns and clinical decision-making differ substantially in patients with kidney failure. Additional VA target populations (Cohorts B-E) were created for sensitivity analyses by relaxing key inclusion criteria. Details of data sources, target population definitions, and covariate measurements are provided in the **Online Supplemental Methods**.

The primary endpoint was time to first major adverse cardiovascular event (MACE), a composite of cardiovascular death, non-fatal myocardial infarction (MI), or non-fatal stroke. Time to non-fatal MI, non-fatal stroke, and all-cause mortality were secondary outcomes. We also evaluated the incidence of selected adverse events. We identified potential effect modifiers and confounders of trial participation: age, self-reported race, measures of diabetes severity (HbA1c, diabetes duration, number of diabetes medications), comorbidities (history of cardiovascular conditions; chronic kidney disease stage; atrial fibrillation; COPD; cancer; liver disease; dementia), and other risk factors (BMI, blood pressure, lipids, kidney function, smoking status). We did not include sex as a covariate in the primary analysis because the VA target population is 97.5% male, creating a near-structural positivity violation; we conducted a sensitivity analysis including sex as a balancing covariate.

Sensitivity Analyses

Several sensitivity analyses were conducted to verify robustness. First, we explicitly included sex as one of the balancing covariates. Second, we performed a leave-one-out analysis to determine whether excluding any single covariate from the set of covariates we balance and adjust in our weighting and outcome models significantly impacted the transported treatment effect estimates. Third, to assess sensitivity to the definition of the target population, we repeated these analyses for alternative cohorts (Cohorts B through E), with particular focus towards populations that include veterans currently ineligible for GLP-1 receptor agonist treatment under existing VA formulary policies. No adjustments for

multiplicity were made given the pre-specified primary outcome (composite MACE) and the exploratory nature of secondary outcome and sensitivity analyses.

Simulation Study

Simulation Design

Data were generated for a combined sample of n = 1,000 individuals partitioned into trial ($Z_i = 1$) and target population ($Z_i = 0$) subsamples with two baseline covariates $X_{i1} \sim Normal(-0.5, 1)$ and $X_{i2} \sim Bernoulli(0.5)$. An unmeasured variable $U_i \sim Normal(0.5, 1)$ was included to induce violations of identifying assumptions in specific scenarios. Trial membership was determined by $Z_i \sim Bernoulli[\pi(X_1, X_2, U)]$, where $\pi(X_{i1}, X_{i2}, U_i) = \text{expit}(\alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 U_i)$, with scenario-specific coefficients $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$. Treatment assignments within the trial followed $A_i \sim Bernoulli(0.5)$, reflecting one-to-one randomization.

Potential event times were generated from a Weibull proportional hazards model with shape parameter $\nu = 1.5$ and baseline hazard $\lambda_0 = 0.2$. Specifically, potential event times were obtained by inverting the Weibull survival function

$$T_i = \left\{ -\frac{\log(V_i)}{\lambda_0 \exp[\eta(A_i, X_{i1}, X_{i2}, U_i)]} \right\}^{1/\nu},$$

where $V_i \sim Uniform(0,1)$ and the linear predictor incorporated treatment, measured covariates, and scenario-specific terms involving U_i: $\eta(A_i, X_{i1}, X_{i2}, U_i) = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 A_i + \beta_4 U_i + \beta_5 A_i \times U_i$. Censoring times C_i were generated independently from an exponential distribution with rate $\exp(-3)$. The observed event time was defined as $\tilde{T}_i = \min(T_i, C_i)$ with event indicator $D_i = I(T_i \leq C_i)$. For individuals in the target population

($Z_i = 0$), treatment assignment and survival outcomes were treated as unobserved, consistent with the transportability framework where only covariate information is available from the target sample.

We considered the survival probability difference between treatment arms as the target estimand at follow-up times $t = 1, 2, 3, 4, 5$, and 6, transported to the target population. Six scenarios were examined: (1) Baseline: both selection and outcome models correctly specified with no unmeasured confounding; (2) Incorrect Outcome Model: the outcome model omits U_i, which influences survival; (3) Incorrect Selection Model: the selection model omits U_i, which influences trial participation; (4) Exchangeability Violation: both models omit U_i; (5) Practical Positivity Violation: extreme coefficients ($\alpha_1 = 2, \alpha_2 = -2$) inducing near-deterministic selection probabilities for some covariate strata; and (6) Structural Positivity Violation: individuals with $X_{i1} > 0$ were deterministically excluded from the trial ($Z_i = 0$).

For the simulation study, the DR estimator was implemented using stable balancing weights,[24] and performance was assessed over 1,000 replications using bias (average estimate minus true parameter value) and 95% confidence interval coverage probability.

Simulation Results

Figure 1 presents bias and coverage probability across follow-up times for each scenario. Under the baseline scenario with correctly specified models, the estimator exhibited negligible bias across all time points (range: −0.008 to 0.003) with coverage probabilities consistently near the nominal 95% level (range: 0.94 to 0.95). These results

demonstrate that the pseudo-observation approach yields valid inference for the transported survival probability difference when underlying model assumptions are satisfied.

The double-robustness property was evident in scenarios involving single-model misspecification. When only the outcome model was misspecified (Scenario 2), bias remained small (range: −0.003 to 0.000) with coverage probabilities between 0.93 and 0.95. When only the selection model was misspecified (Scenario 3), bias was modest (range: −0.003 to 0.006) with coverage between 0.94 and 0.95. These findings confirm that the DR estimator maintains validity when at least one nuisance model is correctly specified.

When both models were misspecified (Scenario 4, exchangeability violation), substantial negative bias emerged (range: −0.039 to −0.049), persisting across follow-up times. Coverage probability deteriorated markedly, falling to 0.81-0.87, reflecting the breakdown of the double-robustness guarantee when neither nuisance model is correctly specified and unmeasured effect modification is present.

Practical positivity violations (Scenario 5) induced modest bias (range: −0.009 to 0.005), though coverage remained acceptable (0.94-0.96). This pattern reflects finite-sample instability from extreme weights rather than fundamental identification failure. Surprisingly, structural positivity violations (Scenario 6) produced small bias (range: −0.005 to 0.006) with coverage near nominal levels (0.94-0.95), suggesting that when model specification remains correct for the observed data, inference remains valid despite deterministic exclusion of certain covariate strata from the trial.

These simulation results support the viability of the pseudo-observation approach for DR transportability estimation under the assumed conditions. The estimator exhibits appropriate bias and coverage when models are correctly specified, and the double-robustness property provides protection against single-model misspecification. However, careful assessment of exchangeability and positivity assumptions remains essential, as violations of either assumption can substantially compromise inference.

Results from Transporting the LEADER Trial

Baseline Characteristics

Cohort A (LEADER-eligible VA cohort) included 357,075 veterans meeting LEADER's inclusion and exclusion criteria, compared to 9,336 participants in the LEADER trial. **Table 1** summarizes baseline characteristics of these two groups. Veterans were older on average (mean age of 70.0 years versus 64.3 years in LEADER), predominantly male (97.5% vs. 64.3% in LEADER), and were more likely to be Black or African American (14.1% vs. 8.3%). Baseline BMI (approximately 32.5 kg/m^2) and HbA1c (8.8% vs. 8.7%) were similar between VA and LEADER samples. Veterans had markedly higher rates of heart failure (21.1% vs. 14.0%), atrial fibrillation (16.3% vs. 1.9%), COPD (22.3% vs. 1.4%), and cancer (36.8% vs. 5.9%) compared to LEADER participants. Veterans also had worse baseline kidney function (mean eGFR 66.4 vs. 79.1 ml/min/1.73m^2) and higher rates of current or former smoking (77.7% vs. 58.6%). After applying sampling weights, SMDs between the LEADER and VA samples were reduced to <0.05 for all balancing variables (**eFigure 1**).

Outcomes and Adverse Events

The transported survival probabilities for the primary MACE outcome and secondary outcomes in the primary VA target population were lower for both the liraglutide and placebo arms than in the original LEADER study (**Figure 2, left**). Randomization to liraglutide was associated with a risk reduction of the primary MACE composite outcome and each secondary outcome compared to placebo. At 36 months, the RD in outcome-free survival favoring liraglutide was 2.01% (95% CI: 0.80%, 3.22%) for composite MACE, 0.17% (−0.05%, 0.85%) for stroke, 1.00% (0.09%, 1.91%) for MI, and 1.50% (0.62%, 2.39%) for all-cause mortality (**Figure 2, right**). All transported RDs fell within the 95% CIs of corresponding LEADER estimates. The RDs of the composite MACE outcome and all-cause mortality, but not for stroke and MI, were qualitatively greater when LEADER was transported to the VA target population.

To assess external validity in the VA target population of safety outcomes assessed in LEADER, we estimated transported effects of liraglutide versus placebo on selected adverse events. As with the primary and secondary efficacy outcomes, the transported RDs for all adverse events examined fell within the 95% confidence intervals estimated in LEADER (**Figure 3**).

Sensitivity Analyses

The transported RDs estimated for the VA target population were nearly identical without and with sex as a balancing variable (**eFigure 2**), supporting our decision to exclude sex from the primary analysis. The minimal difference reflects that balancing on sex in a

97.5% male target population provides little additional adjustment beyond what is achieved by other covariates correlated with sex. In the leave-one-out-analysis, baseline age, HbA1c, and number of diabetes medications qualitatively impacted the transported effect estimates on the absolute (RD) scale for the primary and secondary outcomes (**eFigure 3**). However, regardless of the balancing variable omitted, the transported RD estimated in the VA target population for all outcomes remained within the 95% confidence intervals of the LEADER study treatment effect estimates. Finally, the RD of the primary and secondary efficacy outcomes were similar across all VA target populations – within the 95% confidence intervals of the LEADER study with qualitatively greater absolute risk reduction for the composite MACE and all-cause mortality outcomes (**Figure 4**).

Discussion

Transporting treatment effects across populations, we found that the LEADER trial examining cardiovascular outcomes associated with liraglutide generalized to a real-world cohort of US veterans with type 2 diabetes. Liraglutide was associated with reduced risk of the composite MACE outcome and all-cause mortality in a largely male, elderly VA target population. The transported effect estimates for all outcomes fell within the 95% confidence intervals estimated in the LEADER study, indicating that the cardiovascular benefits observed in LEADER generalize to this veteran population. The point estimates for RD of the primary MACE outcome and all-cause mortality were modestly larger when transported to veterans, corresponding to possibly lower numbers needed to treat (NNT of 50 versus 62.5 for MACE and 67 versus 111 for all-cause mortality over three years).

Moreover, the sensitivity analysis in broader VA target populations suggests liraglutide may confer benefit outside the trial inclusion/exclusion parameters and outside the current criteria for use of GLP-1RA in the VA. This reinforces the cardioprotective role of liraglutide as previously demonstrated in RCTs like LEADER in a real-world population not well-represented in the LEADER RCT.

Our findings extend prior studies that replicate or generalize GLP-1RA cardiovascular outcome trials with real-world data in a variety of settings. The sensitivity analysis results in Cohorts D and E demonstrating cardiovascular benefit of liraglutide in veterans irrespective of pre-existing ASCVD mirror real-world studies evaluating GLP-1RA for ASCVD primary prevention and subgroup analyses in GLP-1RA RCTs that did not demonstrate heterogeneity by ASCVD history.[6,27] Our results also support a recent systematic review that found concordance between RCTs and real-world studies for GLP-1RAs[28] despite real-world patients with diabetes infrequently meeting eligibility criteria for GLP-1RA RCTs.[11] Importantly, transportability analyses explicitly combine RCT data with real-world data unlike prior secondary analyses of RCTs or studies exclusively using real-world data, leveraging the internal validity of the RCT while extending inferences to the target population.

Transportability analyses may help inform treatment policies when constituents of a specific healthcare system are not well-represented in an RCT, and we demonstrate this using data from the largest integrated healthcare system in the US. In the VA, formulary decisions and criteria for use are often based on demonstrated outcomes in veterans. Our study provides evidence that can inform those decisions in the absence of VA-based

RCTs.[29,30] The primary analyses support the use of GLP-1RAs for patients with type 2 diabetes and established ASCVD or high risk of ASCVD. The sensitivity analyses in broader VA target populations suggest that several of the LEADER inclusion criteria may not be essential for liraglutide to be beneficial in Veterans. As several of the RCT inclusion criteria (e.g., prior ASCVD and elevated HbA1c) remain among the criteria for use of GLP-1RA in the VA, these exploratory results with more broadly defined target populations motivate further study of expanding access to these agents, balanced against cost, clinical characteristics, patient preferences, and drug availability. With the recent and ongoing growth in new therapies for cardiometabolic conditions, analogous analyses can guide evaluation of external validity of new treatments in populations relevant to policymakers, payors, and healthcare systems. Critically, transportability analyses can be performed with real-world target population data for potential users of a new medication before treatment uptake in routine care, allowing rapid, time-sensitive real-world evidence generation to inform policies regarding new therapies.

Our study has important limitations. First, transportability assumes that any censoring is non-informative or that the censoring mechanisms would be similar in the target population and in the RCT. Second, while we included many covariates for adjustment, unmeasured differences or effect modifiers between the trial and VA populations could exist. For example, psychosocial factors, diet, exercise, adherence behaviors, or quality of background care might differ and were not captured. However, leave-one-out sensitivity analysis demonstrated that no single effect modifier included in the analysis meaningfully impacted the transported effect estimate, suggesting that the

impact of an unmeasured effect modifier would have to substantially exceed the impact of any of the measured factors examined in the study. Third, the VA population is predominantly older men with extensive comorbidities. While excess disease burden in veterans helps motivate our study, it also means that the external validity of our conclusions is strong for VA and similar target populations, but caution is needed in extrapolating to populations with different characteristics. Fourth, the sensitivity analyses examining broader VA target populations (Cohorts B through E) should be interpreted with caution. By relaxing the inclusion criteria that defined LEADER eligibility — such as age, HbA1c, or ASCVD history — these analyses inherently extrapolate to covariate regions where the trial provides no direct information, representing a structural positivity violation. While the consistent direction of effects across these populations is suggestive, these results cannot definitively establish efficacy outside trial-eligible populations. Rather, they provide hypothesis-generating evidence that the transported effects do not appear to be driven solely by the specific eligibility criteria, motivating further study in broader populations. Finally, even with a very large VA sample, the statistical power for transported estimates is governed by the trial size and the variation in weights. We found that the effective sample size of LEADER that provides information relevant to the primary VA target population was only 2477 out of the 9336 participants in the LEADER study. This means that despite the large LEADER trial, only about one-quarter of participants were similar enough to veterans to contribute meaningful information about treatment effects in the VA population. This effective sample size declined further when examining broader veteran populations in sensitivity analyses (**eFigure 4**). The effective sample size estimates

highlight the limited statistical power to detect significant differences between LEADER and transported effects, though that was not the goal of the study.

By transporting the LEADER trial results to a real-world population, we found that liraglutide's cardiovascular and mortality benefits generalize to US adults with type 2 diabetes receiving care in the VA, the largest integrated healthcare system in the US. The results support the use of liraglutide to mitigate cardiovascular disease risk in a real-world population that was poorly represented in the RCT. The study also illustrates the value of causal transportability analyses in bridging the gap between clinical trial efficacy and real-world effectiveness. As healthcare systems increasingly rely on real-world evidence to inform practice and policy, the approach applied here – leveraging robust RCT results and tailoring them to a target population – can serve as a model for evaluating external validity of RCTs of other new therapies to inform medication access policies.

Acknowledgements

Author contributions

K.J., S.R., W.G.L, and D.G. designed the study. M.M. and K.K. assisted with access to and cleaning of LEADER data for analyses. K.J., W.G.L., and T.W. performed all analyses. K.J. and S.R. drafted the manuscript. All authors provided critical revision and input on the final manuscript.

Funding statement

The funders had no role in the conduct of the study. This publication does not represent the views of the Department of Veteran Affairs or the United States Government. S.R. was supported by VA award IK2-CX001907 and VA award I01-BX006417. J.E.B.R was supported by 1P30DK116073.

Data and code availability statement

Code for all statistical analysis is available upon request. Code to replicate the simulation study can be found at https://github.com/kjosey/. A deidentified, anonymized limited data set derived from the datasets used for the analysis can be made available upon reasonable request from researchers with necessary human subjects research oversight and in accordance with VA data sharing policies.

Conflicts of interest

M.M. and K.K. are employees of Novo Nordisk and provided data access and assistance with reproducing analyses of the original LEADER study. They provided critical feedback on the manuscript but did not shape or modify the reported results, interpretation, or conclusions. The remaining authors report no conflicts of interest.

Transparency

The corresponding authors are the manuscript's guarantors and affirm that the manuscript is an honest, accurate, and transparent account of the study being reported. No important

aspects of the study have been omitted, and there are no discrepancies from the study as originally planned.

Copyright

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide license to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

Table 1. Study participant characteristics at baseline.

	VA	LEADER	p-value
Total N	357,075	9,336	
Male sex, n (%)	348,101 (97.5)	6,001 (64.3)	<0.001
Race, n (%)			<0.001
Black or African-American	50,312 (14.1)	776 (8.3)	
White	279,341 (78.2)	7,237 (77.5)	
Other	27,422 (7.7)	1,323 (14.2)	
Age (years), mean (SD)	70.0 (8.7)	64.3 (7.2)	<0.001
Hemoglobin A1c (%), mean (SD)	8.8 (1.7)	8.7 (1.5)	<0.001
# of diabetes medications, n (%)			<0.001
≤1	173,632 (48.6)	1,313 (14.1)	
2	87,554 (24.5)	3,294 (35.3)	
≥3	95,889 (26.9)	4,729 (50.7)	
BMI (kg/m^2), mean (SD)	32.6 (6.4)	32.5 (6.3)	0.466
eGFR (ml/min/1.73m^2), mean (SD)	66.4 (21.4)	79.1 (22.1)	<0.001
CAD, n (%)	196,030 (54.9)	887 (9.5)	<0.001
HF, n (%)	75,334 (21.1)	1,304 (14.0)	<0.001
Prior stroke, n (%)	85,302 (23.9)	1,506 (16.1)	<0.001
Prior MI, n (%)	44,406 (12.4)	2,862 (30.7)	<0.001
Prior revascularization, n (%)	81,584 (22.8)	3,481 (37.3)	<0.001
CKD stage, n (%)			<0.001
Stage 1	57,932 (16.2)	3,700 (39.6)	
Stage 2	142,760 (40.0)	3,656 (39.2)	
Stage 3 or 4	156,383 (43.8)	1,980 (21.2)	
Any cardiovascular disease, n (%)	303,221 (84.9)	7,731 (82.8)	<0.001
Hypertension, n (%)	306,579 (85.9)	8,508 (91.1)	<0.001
Hyperlipidemia, n (%)	246,848 (69.1)	7,067 (75.7)	<0.001
Atrial fibrillation, n (%)	58,036 (16.3)	179 (1.9)	<0.001
Dementia, n (%)	19,146 (5.4)	16 (0.2)	<0.001
COPD, n (%)	79,735 (22.3)	135 (1.4)	<0.001
Cancer, n (%)	131,325 (36.8)	551 (5.9)	<0.001
Liver disease, n (%)	26,929 (7.5)	594 (6.4)	<0.001
Smoking status, n (%)			<0.001
Current	77,955 (21.8)	1,130 (12.1)	
Former	199,577 (55.9)	4,338 (46.5)	
Never	79,543 (22.3)	3,868 (41.4)	

Abbreviations: BMI, Body mass index; eGFR, estimated glomerular filtration rate; CAD, coronary artery disease; HF, heart failure; MI, myocardial infarction; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease

Figure Legends

Figure 1. Finite-sample performance of the DR transportability estimator across simulation scenarios. Bias (**left panel**) and 95% confidence interval coverage probability (**right panel**) are displayed across follow-up times (t = 1, 2, 3, 4, 5, 6) for six scenarios: Baseline (correct model specification), Incorrect Outcome Model, Incorrect Selection Model, Exchangeability Violation (both models misspecified), Practical Positivity Violation, and Structural Positivity Violation. The dashed horizontal lines indicate zero bias (left) and nominal 95% coverage (right). Results are based on 1,000 simulation replications with n = 1,000 individuals per replication.

Figure 2. (**Left panel**) Predicted survival probabilities in liraglutide (solid lines) and placebo (dashed lines) arms of LEADER trial (blue) and transported to VA target population (yellow) for composite major adverse cardiovascular events (Composite), non-fatal myocardial infarction (MI), non-fatal stroke, and all-cause mortality. (**Right panel**) Absolute risk differences in event-free survival (liraglutide minus placebo) for composite major adverse cardiovascular events (Composite), non-fatal myocardial infarction (MI), non-fatal stroke, and all-cause mortality in LEADER (blue) and transported to VA target population (yellow); 95% confidence intervals are shown with colored shading.

Figure 3. Absolute risk differences at the end of follow-up and 95% confidence intervals for several adverse events recorded in the LEADER trial (blue) and transported to the VA target

population (yellow). Risk differences greater than 0 (vertical dotted line) reflect greater risk of adverse events in individuals randomized to liraglutide compared to placebo.

Figure 4. Comparison of treatment effects in LEADER and LEADER transported to a series of VA target populations with varying inclusion/exclusion criteria. (**A**) Key inclusion criteria for LEADER and for each VA target population included in the analysis. (**B**) Transported treatment effect estimated as survival probability differences for composite major adverse cardiovascular events (Composite), non-fatal myocardial infarction (MI), non-fatal stroke, and all-cause mortality; 95% confidence intervals are shown with colored shading for each transported treatment effect estimate. VA A through VA E indicate results after weighting LEADER to the corresponding target population. Abbreviations: ASCVD, atherosclerotic cardiovascular disease; A1c, hemoglobin A1c; eGFR, estimated glomerular filtration rate calculated using the Modification of Diet in Renal Disease equation (units of ml/min/1.73m^2).

References

1. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet*. 2005;365(9453):82-93.
2. Fahey T. Applying the results of clinical trials to patients to general practice: perceived problems, strengths, assumptions, and challenges for the future. *Br J Gen Pract*. 1998;48(429):1173-1178.
3. US Food and Drug Administration. Diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials guidance for industry. In: Services UDoHaH, ed. Silver Spring, MD2022.
4. Califf RM. Now is the time to fix the evidence generation system. *Clin Trials*. 2023;20(1):3-12.
5. In: Bibbins-Domingo K, Helman A, eds. *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups*. Washington (DC)2022.
6. Badve SV, Bilal A, Lee MMY, et al. Effects of GLP-1 receptor agonists on kidney and cardiovascular disease outcomes: a meta-analysis of randomised controlled trials. *Lancet Diabetes Endocrinol*. 2025;13(1):15-28.
7. Moiz A, Filion KB, Toutounchi H, et al. Efficacy and Safety of Glucagon-Like Peptide-1 Receptor Agonists for Weight Loss Among Adults Without Diabetes : A Systematic Review of Randomized Controlled Trials. *Ann Intern Med*. 2025;178(2):199-217.
8. Knudsen JS, Thomsen RW, Pottegard A, Knop FK, Sorensen HT. Differences Between Randomized Clinical Trial Patients and Real-World Initiators of the Glucagon-Like Peptide 1 Receptor Agonist Liraglutide. *Diabetes Care*. 2018;41(9):e133-e135.
9. Weeda ER, Burnette B, Rowe S, Haas S, Zecopoulos A, Alnewais ME. Disparity in the rural-urban distribution of clinical trial sites used in studies cited by diabetes treatment guidelines. *Diabetes Obes Metab*. 2022;24(3):561-563.
10. Kelley AT, Mizokami-Stout K, O'Brien MJ, Bowen ME, Sussman J. Hispanic representation in diabetes cardiovascular outcomes trials. *BMJ Open Diabetes Res Care*. 2019;7(1):e000656.
11. Wittbrodt ET, Eudicone JM, Bell KF, Enhoffer DM, Latham K, Green JB. Generalizability of glucagon-like peptide-1 receptor agonist cardiovascular outcome trials enrollment criteria to the US type 2 diabetes population. *Am J Manag Care*. 2018;24(8 Suppl):S146-S155.
12. Dugard A, Giraudeau B, Dibao-Dina C. Do patients with type 2 diabetes mellitus included in randomised clinical trials differ from general-practice patients? A cross-sectional comparative study. *BMJ Open*. 2024;14(4):e077582.
13. Sargeant JA, Brady EM, Zaccardi F, et al. Adults with early-onset type 2 diabetes (aged 18-39 years) are severely underrepresented in diabetes clinical research trials. *Diabetologia*. 2020;63(8):1516-1520.
14. Dahabreh IJ, Robertson SE, Steingrimsson JA, Stuart EA, Hernan MA. Extending inferences from a randomized trial to a new target population. *Stat Med*. 2020;39(14):1999-2014.
15. Degtiar I, Rose S. A Review of Generalizability and Transportability. *Annual Review of Statistics and Its Application*. 2023;10(1):null.

16. Josey KP, Berkowitz SA, Ghosh D, Raghavan S. Transporting experimental results with entropy balancing. *Stat Med*. 2021;40(19):4310-4326.
17. Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med*. 2016;375(4):311-322.
18. Marso SP, Bain SC, Consoli A, et al. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. *N Engl J Med*. 2016;375(19):1834-1844.
19. Gerstein HC, Colhoun HM, Dagenais GR, et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. *Lancet*. 2019;394(10193):121-130.
20. Florez HJ, Ghosh A, Pop-Busui R, et al. Differences in complications, cardiovascular risk factor, and diabetes management among participants enrolled at veterans affairs (VA) and non-VA medical centers in the glycemia reduction approaches in diabetes: A comparative effectiveness study (GRADE). *Diabetes Res Clin Pract*. 2022;184:109188.
21. Fryar CD, Herrick K, Afful J, Ogden CL. Cardiovascular Disease Risk Factors Among Male Veterans, U.S., 2009-2012. *Am J Prev Med*. 2016;50(1):101-105.
22. Andersen PK, Perme MP. Pseudo-observations in survival analysis. *Stat Methods Med Res*. 2010;19(1):71-99.
23. Lee D, Yang S, Dong L, Wang X, Zeng D, Cai J. Improving trial generalizability using observational studies. *Biometrics*. 2023;79(2):1213-1225.
24. Keele LJ, Ben-Michael E, Feller A, Kelz R, Miratrix L. Hospital quality risk standardization via approximate balancing weights. *The Annals of Applied Statistics*. 2023;17(2):901-928, 928.
25. Josey KP, Juarez-Colunga E, Yang F, Ghosh D. A framework for covariate balance using Bregman distances. *Scandinavian Journal of Statistics*. 2021;48(3):790-816.
26. van der Laan MJ, Polley EC, Hubbard AE. Super Learner. *Statistical Applications in Genetics and Molecular Biology*. 2007;6(1).
27. Richardson TL, Jr., Halvorson AE, Hackstadt AJ, et al. Primary Occurrence of Cardiovascular Events After Adding Sodium-Glucose Cotransporter-2 Inhibitors or Glucagon-like Peptide-1 Receptor Agonists Compared With Dipeptidyl Peptidase-4 Inhibitors: A Cohort Study in Veterans With Diabetes. *Ann Intern Med*. 2023;176(6):751-760.
28. Wang W, Tang W-C, Webster-Clark M, Yu OH, Filion KB. Replicating cardiovascular outcome trials of medications used to treat type 2 diabetes using real-world data: A systematic review of observational studies. *medRxiv*. 2024:2024.2011.2028.24317099.
29. Tummalapalli SL, Montealegre JL, Warnock N, Green M, Ibrahim SA, Estrella MM. Coverage, Formulary Restrictions, and Affordability of Sodium-Glucose Cotransporter 2 Inhibitors by US Insurance Plan Types. *JAMA Health Forum*. 2021;2(12):e214205.
30. Zerzan JT, Gibson M, Libby AM. Improving state Medicaid policies with comparative effectiveness research: a key role for academic health centers. *Acad Med*. 2011;86(6):695-700.

Figure 1.

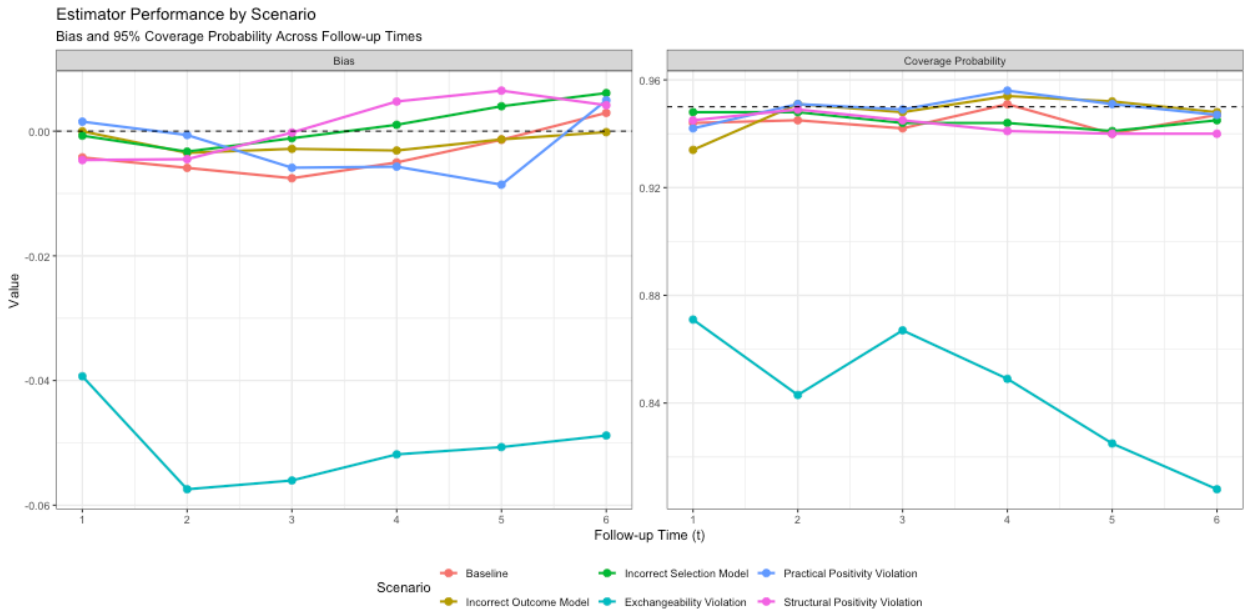

Estimator Performance by Scenario
Bias and 95% Coverage Probability Across Follow-up Times
Bias
Coverage Probability
Value
0.00
-0.02
-0.04
-0.06
0.96
0.92
0.88
0.84
1
2
3
4
5
6
Follow-up Time (t)
Scenario
Baseline
Incorrect Outcome Model
Incorrect Selection Model
Exchangeability Violation
Practical Positivity Violation
Structural Positivity Violation

Figure 2.

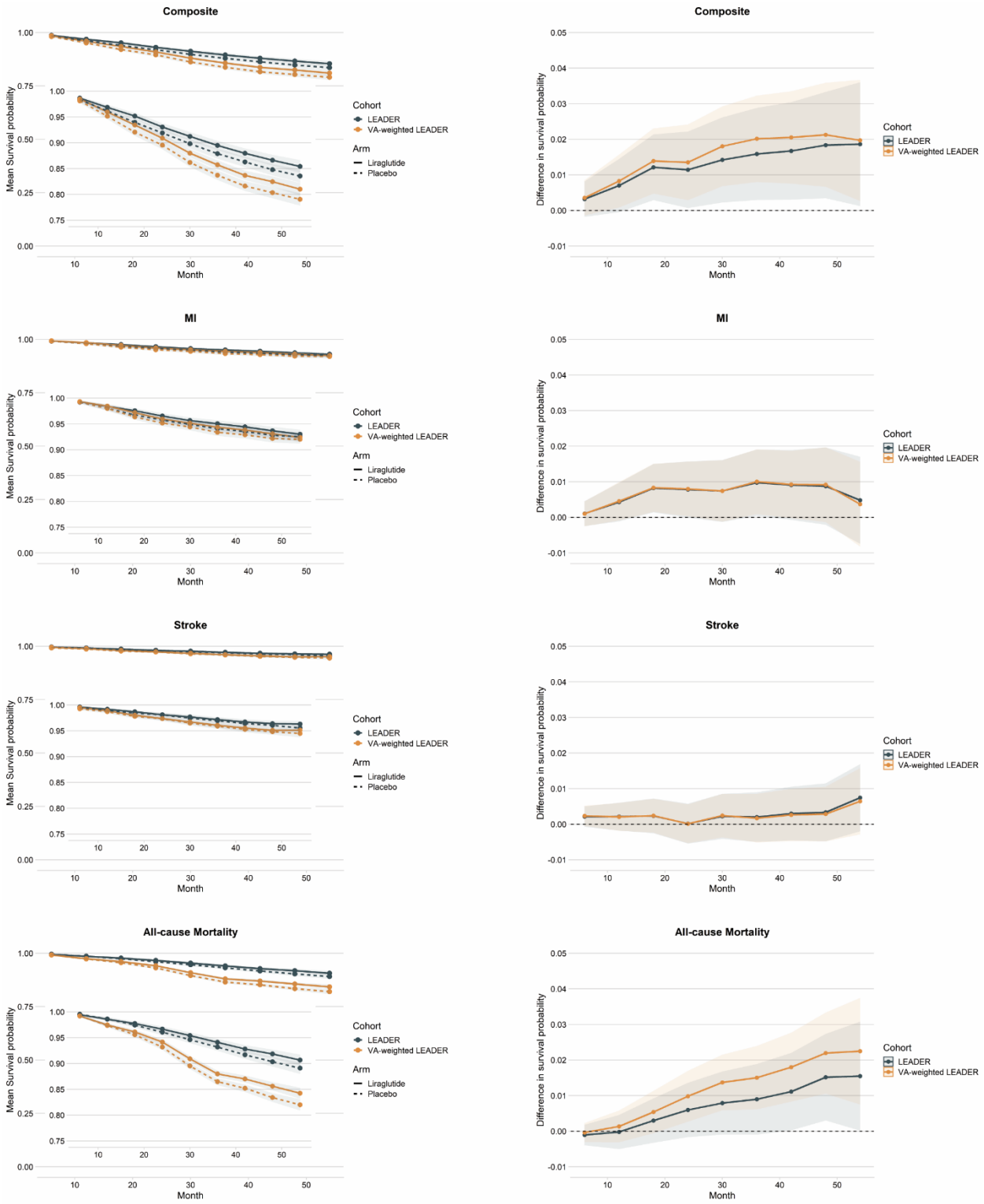

Composite
MI
Stroke
All-cause Mortality
Mean Survival probability
Difference in survival probability
Month
Cohort
LEADER
VA-weighted LEADER
Arm
Liraglutide
Placebo

Figure 3.

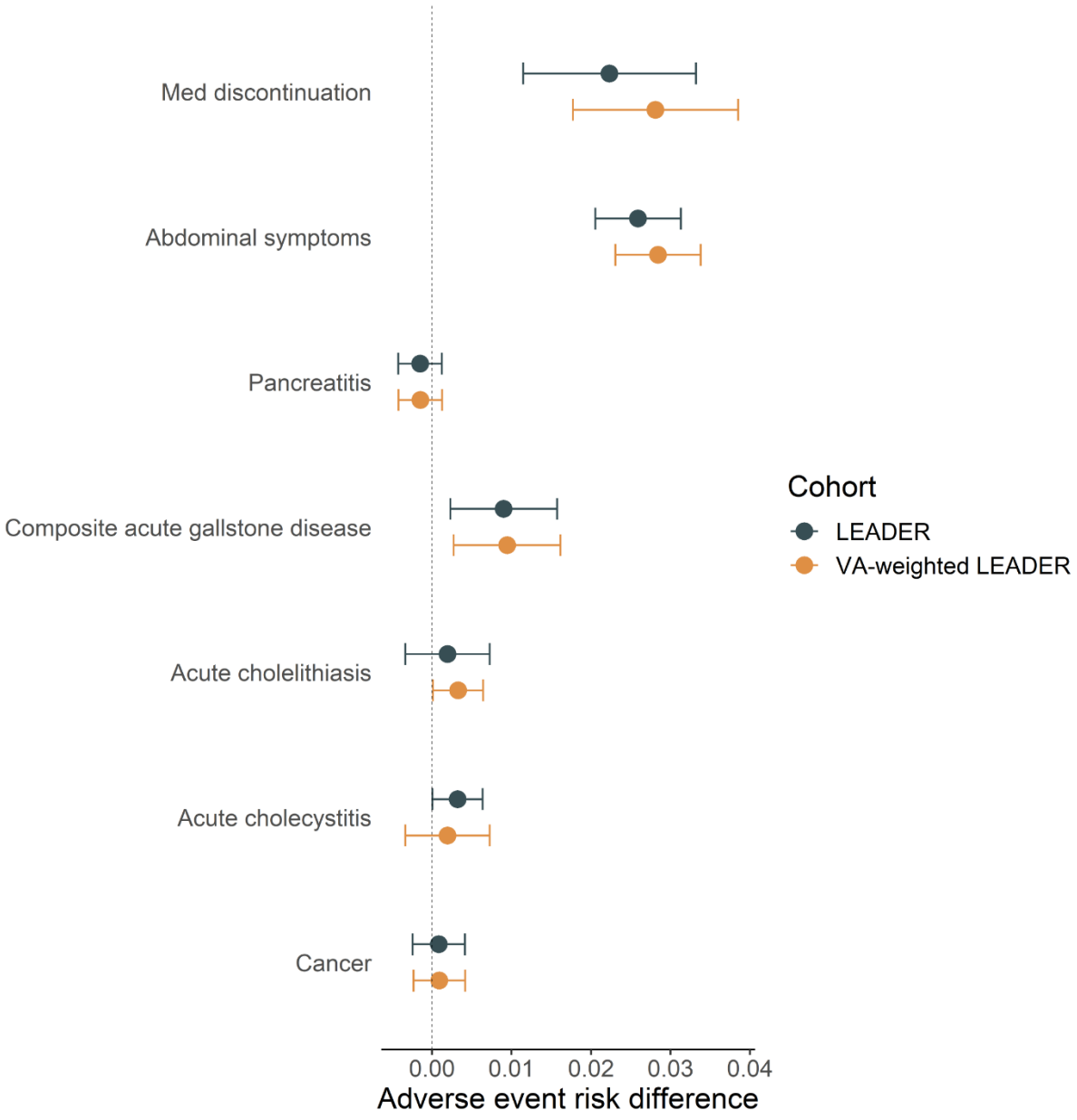
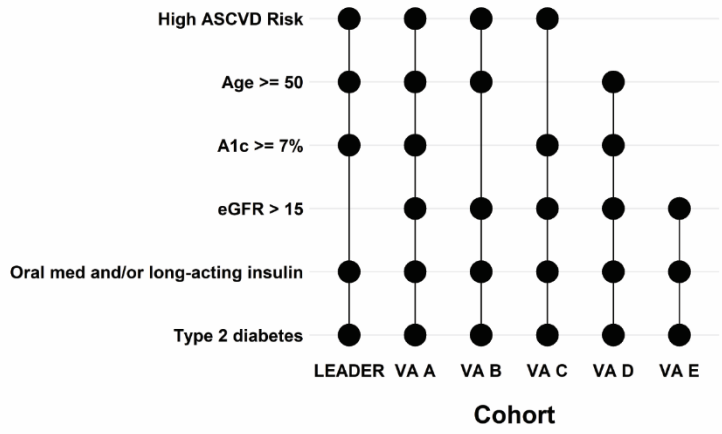
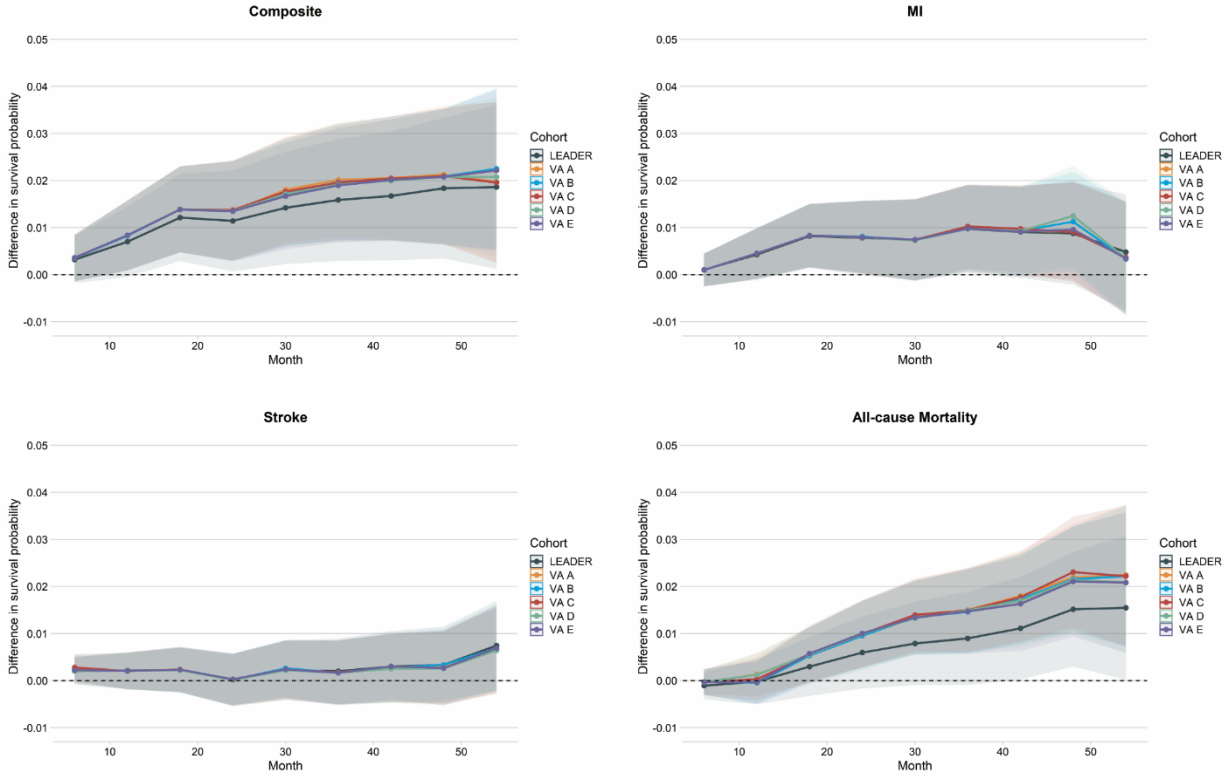

Med discontinuation
Abdominal symptoms
Pancreatitis
Composite acute gallstone disease
Acute cholelithiasis
Acute cholecystitis
Cancer
0.00
0.01
0.02
0.03
0.04
Adverse event risk difference
Cohort
LEADER
VA-weighted LEADER

Figure 4.

A.


High ASCVD Risk
Age >= 50
A1c >= 7%
eGFR > 15
Oral med and/or long-acting insulin
Type 2 diabetes
LEADER
VA A
VA B
VA C
VA D
VA E
Cohort

B.


Composite
MI
Stroke
All-cause Mortality
Difference in survival probability
Month
Cohort
LEADER
VA A
VA B
VA C
VA D
VA E