Enrolment criteria for diabetes cardiovascular outcome trials do not inform on generalizability to clinical practice. The case of GLP-1 receptor agonists

Veronica Sciannameo[1], Paola Berchialla[2], Emanuela Orsi[3], Olga Lamacchia[4], Susanna Morano[5], Fabrizio Querci[6], Agostino Consoli[7], Angelo Avogaro[8], Gian Paolo Fadini[8]
on behalf of the DARWIN-T2D study*

[1] Unit of Biostatistics, Epidemiology and Public Health. Department of Cardiac, Thoracic, Vascular Sciences and Public Health. University of Padova, 35128 Padova, Italy.
[2] Department of Clinical and Biological Sciences, University of Turin, 10043 Turin, Italy.
[3] Unit of Endocrinology and Metabolic Diseases, Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico di Milano, 20122 Milan, Italy.
[4] Unit of Endocrinology, Department of Medical and Surgical Sciences, University Hospital of Foggia, 71122 Foggia, Italy.
[5] Unit of Diabetes Complications, V Clinica Medica, Department of Experimental Medicine, University of Rome "La Sapienza", 00161 Rome, Italy.
[6] Unit of Diabetology, ASST Bergamo Est, 24022 Alzano Lombardo, Italy.
[7] Department of Medicine and Aging Science, "G. D'Annunzio" University of Chieti, 68100 Chieti, Italy.
[8] Department of Medicine, University of Padova, 35128 Padova, Italy.
* See Appendix.

Running title: Generalizability of GLP-1RA cardiovascular outcome trials

Correspondence
Gian Paolo Fadini, MD PhD
Associate Professor of Endocrinology
Department of Medicine, University of Padova
Via Giustiniani 2, 35128 Padova, Italy.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/dom.13962

Phone: +39 049 8214318
Fax: +39 049 8212184
Email: gianpaolofadini@hotmail.com

Abstract

Aims. To evaluate generalizability of cardiovascular outcome trials (CVOTs) on GLP-1 receptor agonists (GLP-1RA), we assessed which proportion of real-world patients with of type 2 diabetes (T2D) constitute true CVOT-like populations.

Materials and methods. We applied inclusion/exclusion (I/E) criteria of each GLP-1RA CVOT on a cross-sectional database of 281,380 T2D patients from Italian diabetes outpatient clinics. We calculated the proportion of patients eligible for each CVOT and compared their clinical characteristics with those of trial patients. In addition, we used a Bayesian network-based method to sample the greatest subsets of real-world patients yielding true CVOT-like populations.

Results. Between 98,725 and 124,164 T2D patients could be evaluated for CVOT eligibility. After excluding patients who already were on GLP-1RA and applying I/E criteria, 35.8% of patients would be eligible for REWIND, 34.1% for PIONEER-6, 13.4% for EXSCEL, 10.1% for SUSTAIN-6, 9.5% for HARMONY and 9.4% for LEADER. 45.4% of patients could be eligible for at least one of the CVOTs. These patients, however, were extremely different from trial patients for most clinical characteristics, including demographics, concomitant medications, and complications. The greatest CVOT-like subset of real-world patients was 0.5% for SUSTAIN-6, 1.0% for EXSCEL, 1.2% for LEADER, 1.8% for PIONEER-6, and 7.9% for REWIND.

Conclusions. A very small proportion of real-world patients constitute true CVOT-like populations. These findings question whether any meaningful information can be drawn from applying trial enrolment criteria to real-world T2D patients.

Introduction

In the field of diabetes pharmacotherapy, cardiovascular outcome trials (CVOTs) have been designed to address cardiovascular safety of glucose lowering medications (GLMs) against placebo [1]. Although designed primarily to demonstrate non-inferiority, some CVOTs have shown that active treatment was superior to placebo in reducing the rates of major adverse cardiovascular outcome events (MACE) in patients with type 2 diabetes (T2D). Results of these CVOTs have been incorporated into consensus algorithms, which now prioritize certain GLM for the prevention of cardiovascular complications [2].

This is the case for glucagon-like peptide-1 receptor agonists (GLP-1RA) and sodium glucose contransporter-2 inhibitors (SGLT-2i). Among GLP-1RA, dedicated CVOTs showed superiority of liraglutide (LEADER), semaglutide (SUSTAIN-6), albiglutide (HARMONY), and dulaglutide (REWIND) versus placebo in reducing the rates of 3-point MACE (cardiovascular death, non-fatal myocardial infarction or stroke) [3-6]. The CVOT on exenatide once weekly (exeOW) and oral semaglutide showed nominal reductions in mortality rates [7, 8]. Therefore, there is general agreement that GLP-1RA as a class improve cardiovascular outcomes of T2D [9].

However, generalizability of CVOT findings to clinical practice may be hampered by the many differences between the trial setting and routine care. CVOTs recruited patients based on rigorous inclusion / exclusion (I/E) criteria and closely followed them at regular intervals within strict trial experimental protocols. CVOTs are event-driven, and their inclusion/exclusion (I/E) criteria allow rapid collection of the desired number of events by selecting T2D patients with a prior history of cardiovascular disease and/or multiple risk factors. Since only about 30% of T2D patients in routine care show signs of macroangiopathy, to what extent results of CVOTs can be translated to T2D patients at lower cardiovascular risk is a matter of debate. Positive results of the REWIND, wherein ~30% of patients had established cardiovascular disease [3], suggest that GLP-1RA benefit may be independent from the baseline cardiovascular risk [9, 10].

A few studies have examined what proportion of T2D patients from various clinical care settings would satisfy I/E criteria to be enrolled in specific CVOTs [11-14], but no study so far examined which is the proportion of real-world patients that corresponds to CVOT populations. Indeed, applying CVOT enrolment criteria to real-world T2D patients yields subgroups that are substantially different from the actual CVOT populations, thereby leaving the question of generalizability unanswered.

In this study, we aimed to resolve this issue. After applying I/E criteria for GLP-1RA CVOTs to a T2D real-world population, we show how much the eligible population of patients differs from those of CVOTs and calculate what proportion of patients from routine care would generate a true CVOT-like population.

Methods

Data source. The DARWIN-T2D (DAta for Real World evIdeNce in Type 2 Diabetes) study was conducted by the Italian Diabetes Society and was initially designed to evaluate dapagliflozin in the real world. Protocol details and primary analysis have been published before [15, 16]. The DARWIN-T2D database contains cross-sectional information on about 281k T2D patients from 46 diabetes outpatient clinics from all over Italy. Data were collected at each center at the last available visit attended by each patient in years 2015 and 2016. All clinics used the same electronic chart system to store patients' data (MyStar Connect / Smart Digital Clinic, Meteda Srl, San Benedetto del Tronto, Italy). Relevant data were extracted by a dedicated software without manual intervention.

We recorded the following information: demographics (age, sex, diabetes duration), anthropometrics (height, weight, BMI, waist circumference), cardiovascular risk factors (smoke, blood pressure values, lipid profile), estimated glomerular filtration rate (CKD-EPI equation [17]) and other laboratory data (including urinary albumin excretion rate and liver enzymes), and medications for the treatment of diabetes and other cardiovascular risk factors or conditions. In addition, detailed information were collected on diabetic complications from ICD-9 codes in the electronic chart, including: presence and stage of retinopathy and diabetic macular oedema; presence or absence of somatic or autonomic diabetic neuropathy; history of stroke, transient ischemic attack, or carotid endoarterectomy / stenting; history of angina, myocardial infarction, or coronary revascularization; presence / absence of left ventricular hypertrophy and history of heart failure; history of claudication, limb ischemia or amputation; presence of asymptomatic atherosclerosis of coronary, carotid, or leg arteries. Codes to define prevalent cardiovascular disease have been standardized in the electronic chart, which is routinely used to monitor quality of diabetes care in Italy [18].

Not all records were complete for all patients: degree and distribution of missing variables has been shown before [15].

Data analysis. We retrieved, from the respective publications, I/E criteria of the following CVOTs on GLP-1RA: LEADER (liraglutide) [6], SUSTAIN-6 (semaglutide) [5], EXSCEL (exeOW) [7], REWIND (dulaglutide) [3], PIONEER-6 (oral semaglutide) [8], HARMONY (albglutide) [4]. Such trials have been published from 2015 to 2019, a timeframe partially overlapping with our data collection period (2015-2016). Specific I/E criteria had to be adapted to the available data in the DARWIN-T2D database with some modifications (Table S1). For instance, the following information used for CVOT I/E criteria were not

available: myocardial ischemia stress test or imaging; diastolic dysfunction; ankle-brachial index; calcitonin levels, history of cancer and pancreatitis. Timing of prior cardiovascular events and revascularization was not available to exclude patients with recent events. Since no information on contraception was available, women of childbearing potential were excluded. Patients with missing information for key I/E variables were excluded from the analysis.

We thus identified T2D patients who would be eligible into each of the considered CVOTs and calculated the respective proportion over the total background population of patients with available data. Then, we compared the average clinical characteristics of eligible patients with average clinical characteristics of patients in the database who were already on GLP-1RA and with those of patients actually enrolled in CVOTs (from the respective publications).

Finally, we extracted from the DARWIN-T2D database the largest subgroup of patients who had average clinical characteristics superimposable to those of CVOTs. Since, no specific tool is available for this case sampling procedure, we devised an analytical strategy as described below.

Statistical analysis. For descriptive purposes, continuous variables are expressed as mean and standard deviation, whereas categorical variables are expressed as percentage. To evaluate to what extent two groups of patients were similar, we computed the absolute standardized mean difference (SMD) for each variable. Conventionally, a SMD value of 0.1 or less is considered indicative of a good balance. For example, for continuous variables, a SMD <0.1 means that the difference between means of the two groups is <10% of the pooled standard deviation. Due to the very large sample size in each comparison, p-values were not calculated, as several minor and clinically-irrelevant differences would yield p-values <0.05.

Sampling CVOT-like populations. Continuous variables in the DARWIN-T2D database were categorized into five classes, on the basis of each trial summary statistics and assuming a normal distribution. Observations with at least one missing value were removed. A Bayesian network (BN) was constructed for each RCT, because each variable was pre-processed (continuous variables were categorized) according to the specific RCT summary statistics, on the case complete dataset to obtain the conditional probability distributions, which reflect the dependencies among variables. All the available variables were included in the construction of the BN. Then, the PC (Peter-Clark) stable algorithm with a 100-fold bootstrap was employed for the structural learning of the BN (i.e. for identifying the relationships among variables) [19]. A more robust BN was obtained by averaging the 100 BNs learned, considering only the relationships among variables present in at least 95% of times [20]. Bootstrap was performed to address the sample variability. Finally, a Maximum a Posteriori

estimation was used to compute the set of probabilities for conditional nodes, whereas for unconditional nodes, probabilities were assigned computing the ratio between CVOT frequency and DARWIN-T2D frequency, for each variable category, and then normalizing to 1. This was done to take into account the different distribution of patients' characteristics between DARWIN-T2D and CVOT and to increase the probability of being selected of patients less represented in DARWIN-T2D than CVOT and vice-versa to decrease the probability of selection of patients more represented in DARWIN-T2D than CVOT. The decomposition of the join probability into the product of conditional and unconditional probabilities was used to get a final probability of inclusion in CVOT for each patients in DARWIN-T2D [21]. Balancing between the patients sampled and the trial's patients was evaluated through SMD. To get SMD smaller than 0.1 we proceeded in the following way: first, we balanced the two groups according to a SMD smaller than 0.2 for each variable. To obtain this result, for each variable with SMD greater than 0.2, 2% of patients with values in the tails of the distribution were sampled and removed from the DARWIN-T2D dataset. This procedure was iterated until SMD was smaller than 0.2.

The whole procedure was repeated to get a new random sample of patients balanced according to SMD smaller than 0.2. Finally, all the balanced groups obtained were joint together, and again to obtain for each variable SMD smaller than 0.1 the same procedure was applied. A sensitivity analysis on different thresholds of SMD was carried out, and the choice of the double threshold 0.2 and 0.1 turned out to get the greatest balanced group. All the analyses were performed using R version 3.5.0.

Results

The original dataset was composed of 281,380 patients. Since the minimum requirement to enter the database s T2D diagnosis, many patients had missing values for several of the variables needed to evaluate CVOT I/E criteria. Among 130,380 patients with available information on GLM, 6699 (5.1%) were being treated with a GLP-1RA (73.8% liraglutide; 23.5% exeOW; 2.7% lixisenatide). The number of patients who could be evaluated for CVOT eligibility was 124,164 for EXSCEL, 116,553 for PIONEER-6, 107,040 for HARMONY, 106,606 for LEADER, 105,074 for REWIND, and 98,725 for SUSTAIN-6.

Patients already taking GLP-1RA at the visit being examined were considered separately, because ongoing GLP-1RA therapy was an exclusion criterion in CVOTs. After applying I/E criteria as outlined in Table S1, we calculated that the percentage of patients who would be eligible for CVOTs was 35.8% for REWIND,

34.1% for PIONEER-6, 13.4% for EXSCEL, 10.1% for SUSTAIN-6, 9.5% for HARMONY and 9.4% for LEADER. 45.4% of patients could be eligible for at least one of the CVOTs considered.

Clinical characteristics of patients treated with GLP-1RA and of those who could be eligible for CVOTs are illustrated in Table 1.

We observed that the average clinical characteristics of patients who could be eligible for CVOTs were substantially different from the average clinical characteristics of patients who composed each CVOT population (Figure 1A). For instance, patients selected from the real-world database were older than in CVOTs. In addition, despite 80-100% of patients in the LEADER and SUSTAIN-6 had established cardiovascular disease, application of I/E criteria to the real-world population yielded patients with a 70-80% prevalence of microangiopathy (mostly chronic kidney disease) and a lower prevalence of macroangiopathy (40-50%). Most other clinical characteristics were imbalanced between patients enrolled in CVOTs and real-world patients eligible for the same CVOTs (Table 2). Out of 11 key clinical variables, eligible patients matched trial characteristics with absolute SMD <0.1 for just 2 or 3 variables, with the notable exception of REWIND. Real-world patients eligible for REWIND were matched with the REWIND population for 6/11 variables.

Patients who already were on GLP-1RA showed substantially different clinical characteristics when compared to both those satisfying CVOT I/E criteria and those actually enrolled in CVOTs.

We then evaluated which proportion of real-world patients would constitute a population of individuals with key average characteristics similar to those enrolled in CVOTs. The largest dataset of real-world patients yielding CVOT-like populations was 0.5% for SUSTAIN-6, 1.0% for EXSCEL, 1.2% for LEADER, 1.8% for PIONEER-6, and 7.9% for REWIND. We were unable to obtain a meaningful dataset of real-word patients who would match the population of the HARMONY study (Figure 1B).

Discussion

Although 10-35% of real-world T2D patients could be enrolled in GLP-1RA CVOTs, their clinical characteristics were markedly different from those of CVOT populations. For the first time, we calculated that the proportion of real-world patients who have true CVOT-like characteristics is much smaller, ranging from 0.5% to 7.9%.

CVOTs have shown notable capacity of some GLP-1RA to reduce the rate of adverse cardiovascular outcomes in patients with T2D [9]. Generalizability of such findings to clinical practice is challenging especially because the populations investigated in CVOTs are markedly different from real-world T2D patients. CVOTs select patients whose characteristics are intended to maximize the probability of trial success, whereas representativeness of the real-world population of T2D is rarely an issue considered in CVOT design.

Prior studies have examined what proportion of patients from clinical practice databases would be eligible for CVOTs on GLP-1RA or SGLT-2i. By analysing U.S. adult T2D databases, Boye et al. reported proportions of patients eligible for the LEADER, SUSTAIN-6, EXSCEL, and REWIND that were quite similar to those shown in our study [13]. This comparison is important because most CVOTs were conducted in the U.S. and the proportion of patients eligible for CVOTs could be expected to be higher in U.S. than in European databases. While small differences were probably due to regional variations (e.g. in BMI), no major difference emerged, suggesting a limited impact of geographical and cultural factors.

Similar analyses have been performed on CVOTs for SGLT-2i [12, 14]. Nicolucci et al. showed that real-world T2D patients eligible for CVOTs on SGLT-2i were different from trial populations in many instances [11]. So far, however, no study was able to calculate which proportion of real-world patients truly constitute CVOT-like populations.

By analysing GLP-1RA CVOTs, we found substantial differences between the eligible real-world populations and trial populations. The high proportion of patients eligible for PIONEER-6 [8] in our study reflects enrolment criteria that, differently from those of EXSCEL [7], lacked constrain on the ratio between patients with established cardiovascular disease or multiple cardiovascular risk factors. Yet, the resulting PIONEER-6 eligible subset was greatly imbalanced compared to the true PIONEER-6 population [8]. We thus examined which proportion of patients from the real-world database would generate CVOT-like populations. To this end, we used a Bayesian method to sample patients from a large dataset based on given average clinical characteristics. We found that the greatest subset of patients with CVOT-like characteristics was much smaller than the proportion of eligible patients. Interestingly, we found no subset of real-world patients matching the HARMONY trial population, likely because HARMONY patients had established cardiovascular disease at a relatively young age [4]. Notably, REWIND confirmed as the CVOT mostly represented within the T2D population, although only 7.9% of real-world patients were truly REWIND-like. On the contrary, the apparently large generalizability of PIONEER-6 based on I/E criteria was not confirmed. This important finding highlights that CVOT populations are extremely specific and that they are poorly represented by real-world T2D patients.

We acknowledge that, in view of the potentially wide cardiovascular benefits of GLP-1RA, this class of GLM is far underutilized among T2D patients [22]. However, our data argue that generalizability of trial populations to clinical practice should not be based on trial I/E criteria. It is important to remind that representativeness of trial populations is only one aspect of trial finding generalizability. Observational studies, although unable to substitute for CVOTs [23], complement CVOT findings and can provide an estimate of true effectiveness in real-world patients often not represented in CVOTs. Real-world studies on glycaemic and extra-glycaemic

effectiveness of GLP-1RA have confirmed findings from phase III RCTs [24-26], but there is still a relative paucity of cardiovascular real-world studies [27]. In a small study from U.K., intensification of oral therapy by adding GLP-1RA was associated with lower cardiovascular events rate than intensification with insulin [28]. A study using the Swedish and Danish diabetes registries reported that liraglutide, compared to DPP-4 inhibitors, exerted significant cardiovascular protection only in patients with established cardiovascular disease [29]. In parallel to observational studies, alternative methods have been recently developed to transport trial findings to target populations using an inverse odds weighting approach [30, 31].

Our study has limitations inherent to the nature of the data being analysed. Real-world databases often lack some of the information required to address trial I/E criteria, leading to the need of simplifications (Table S1). Exclusion of patients with missing data can lead to deviations from the target population, potentially affecting generalizability analysis [32]. Under-reporting may be another issue in data collected for clinical purposes, possibly leading to an underestimation of the proportion of patients eligible for CVOTs. Finally, we excluded prevalent GLP-1RA users from the analysis of generalizability because it is expected GLP-1RA would have already determined at least part of their benefits in those patients. However, clinical characteristics of the small group of patients on GLP-1RA did not resemble those of CVOT populations, suggesting that excluding prevalent GLP-1R users was unlikely to affect generalizability findings in a meaningful manner.

In conclusion, we show that applying CVOT enrolment criteria to real-world populations leads to a huge overestimation of patients who resemble CVOT populations, thereby leaving open the issue of generalizability.

Acknowledgements

We wish to thank Alessia Russo, secretariat of the Italian Diabetes Society, for the invaluable technical support.

Funding

The study was supported by the Italian Diabetes Society and the University of Padova.

Conflict of interest disclosure

GPF received grant support, lecture or advisory board fees from AstraZeneca, Boehringer-Ingelheim, Eli Lilly, Mundipharma, NovoNordisk, Sanofi, Genzyme, Servier, Abbott, Novartis, Merck Sharp & Dohme. EO received lecture fees from Eli lilly and Novo Nordisk. OL has received speaker fees from Eli-Lilly, NovoNordisk. She also Received Research Grants From Astra zeneca and advisory board fees from NovoNordisk. SM received lecture or advisory board fees from AstraZeneca, Sanofi, Takeda. AC has received

consultancy or speaker fees from Abbott, AstraZeneca, Boehringer-Ingelheim, Bruno Farmaceutici, Janssen, Eli-Lilly, Merck Sharp & Dohme, Novartis, NovoNordisk, Roche, Sanofi-Aventis, Servier, Takeda. He Also Received Research Grants From Eli-Lilly and NovoNordisk. AA received research grants, lecture or advisory board fees from Merck Sharp & Dome, AstraZeneca, Novartis, Boeringher-Ingelheim, Sanofi, Mediolanum, Janssen, NovoNordisk. VS, PB and FQ have nothing to disclose.

Author contribution

Study design: VS, PB, AC, AA, GPF. Data collection and analysis: VS, PB, EO, OL, SM, FQ, GPF. Manuscript writing: VS, PB, AC, AA, GPF. Manuscript revision: EO, OL, SM, FQ. All authors approved the final version of the manuscript.

Composition of the DARWIN-T2D database

Agostino Consoli and Gloria Formoso (Dipartimento di Medicina e Scienze dell'Invecchiamento - Università Degli studi G. D'Annunzio di Chieti-Pescara); Giovanni Grossi (Ospedale San Francesco di Paola - Azienda Sanitaria Provinciale di Cosenza); Achiropita Pucci (Azienda Sanitaria Provinciale di Cosenza); Giorgio Sesti and Francesco Andreozzi (Azienda Ospedaliero Universitaria di Catanzaro); Giuseppe Capobianco (Azienda Sanitaria Locale Napoli 2 Nord); Adriano Gatti (Ospedale San Gennaro dei Poveri - Azienda Sanitaria Locale Napoli 1 Centro); Riccardo Bonadonna, Ivana Zavaroni and Alessandra Dei Cas (Azienda Ospedaliero Universitaria di Parma); Giuseppe Felace (Ospedale di Spilimbergo - Azienda per l'Assistenza Sanitaria n.5 Friuli Occidentale); Patrizia Li Volsi (Ospedale di Pordenone - Azienda per l'Assistenza Sanitaria n.5 Friuli Occidentale); Raffaella Buzzetti and Gaetano Leto (Ospedale Santa Maria Goretti - Azienda Sanitaria Locale di Latina); Gian Pio Sorice (Fondazione Policlinico Universitario A. Gemelli, Roma); Paola D'Angelo (Ospedale Sandro Pertini - Azienda Sanitaria Locale Roma 2); Susanna Morano (Azienda Ospedaliera Universitaria Policlinico Umberto I, Roma); Antonio Carlo Bossi (Ospedale di Treviglio - Azienda Socio Sanitaria Territoriale Bergamo Ovest); Edoardo Duratorre (Ospedale Luini Confalonieri di Luino - Azienda Socio Sanitaria Territoriale Sette Laghi); Ivano Franzetti (Ospedale Sant'Antonio Abate di Gallarate - Azienda Socio Sanitaria Territoriale Valle Olona); Paola Silvia Morpurgo (Ospedale Fatebenefratelli - Azienda Socio Sanitaria Territoriale Fatebenefratelli Sacco); Emanuela Orsi (Fondazione IRCCS Ca' Granda - Ospedale Maggiore Policlinico di Milano); Fabrizio Querci (Ospedale Pesenti Fenaroli di Alzano Lombardo - Azienda Socio Sanitaria Territoriale Bergamo Est); Massimo Boemi† and Federica D'Angelo (Presidio Ospedaliero di Ricerca INRCA-IRCCS di Ancona); Massimiliano Petrelli (Azienda Ospedaliero Universitaria Ospedali Riuniti di Ancona); Gianluca Aimaretti and Ioannis Karamouzis (Azienda Ospedaliero Universitaria Maggiore

della Carità di Novara); Franco Cavalot (Azienda Ospedaliero Universitaria San Luigi Gonzaga, Orbassano); Giuseppe Saglietti† (Ospedale Madonna del Popolo di Omegna - Azienda Sanitaria Locale Verbano Cusio Ossola); Giuliana Cazzetta (Casa della Salute, Ugento - Distretto Socio Sanitario Gagliano del Capo - Azienda Sanitaria Locale di Lecce); Silvestre Cervone (Presidio ospedaliero San Marco in Lamis - Distretto Socio Sanitario San Marco in Lamis - Azienda Sanitaria Locale di Foggia); Eleonora Devangelio (Distretto Socio Sanitario di Massafra - Azienda Sanitaria Locale di Taranto); Olga Lamacchia (Azienda Ospedaliero Universitaria Ospedali Riuniti di Foggia); Salvatore Arena (Ospedale Umberto I – Azienda Sanitaria Provinciale di Siracusa); Antonino Di Benedetto (Azienda Ospedaliera Universitaria Policlinico G. Martino di Messina); Lucia Frittitta (Azienda Ospedaliera di Rilievo Nazionale e di Alta Specializzazione Garibaldi di Catania); Carla Giordano (Azienda Universitaria Policlinico Paolo Giaccone di Palermo); Salvatore Piro (Azienda Ospedaliera di Rilievo Nazionale e di Alta Specializzazione Garibaldi di Catania); Manfredi Rizzo, Roberta Chianetta and Carlo Mannina (Azienda Universitaria Policlinico Paolo Giaccone di Palermo); Roberto Anichini (Ospedale San Jacopo di Pistoia – Azienda USL Toscana Centro); Giuseppe Penno (Azienda Ospedaliero Universitaria Pisana); Anna Solini (Azienda Ospedaliera Universitaria Pisana); Bruno Fattor (Comprensorio Sanitario di Bolzano - Azienda Sanitaria della Provincia Autonoma di Bolzano); Enzo Bonora and Massimo Cigolini (Azienda Ospedaliero Universitaria Integrata di Verona); Annunziata Lapolla and Nino Cristiano Chilelli (Complesso Socio Sanitario Ai Colli - Azienda ULSS n.6 Euganea); Maurizio Poli (Ospedale Girolamo Fracastoro di San Bonifacio - Azienda ULSS n.9 Scaligera); Natalino Simioni and Vera Frison (Ospedale di Cittadella - Azienda ULSS n.6 Euganea); Carmela Vinci (Azienda ULSS n.4 Veneto Orientale).

References

[1] Kieffer CM, Robertson AS. Impact of FDA-Required Cardiovascular Outcome Trials on Type 2 Diabetes Clinical Study Initiation From 2008 to 2017. Ther Innov Regul Sci. 2019: 2168479019860122

[2] Davies MJ, D'Alessio DA, Fradkin J, *et al.* Management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia. 2018; **61**: 2461-2498

[3] 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**: 121-130

[4] Hernandez AF, Green JB, Janmohamed S, *et al.* Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. Lancet. 2018; **392**: 1519-1529

[5] Marso SP, Bain SC, Consoli A, *et al.* Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. N Engl J Med. 2016; **375**: 1834-1844

[6] Marso SP, Daniels GH, Brown-Frandsen K, *et al.* Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med. 2016; **375**: 311-322

[7] Holman RR, Bethel MA, Mentz RJ, *et al.* Effects of Once-Weekly Exenatide on Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med. 2017; **377**: 1228-1239

[8] Husain M, Birkenfeld AL, Donsmark M, *et al.* Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. N Engl J Med. 2019; **381**: 841-851

[9] Kristensen SL, Rorth R, Jhund PS, *et al.* Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. Lancet Diabetes Endocrinol. 2019; **7**: 776-785

[10] Mannucci E, Dicembrini I, Nreu B, Monami M. Glucagon-like peptide-1 receptor agonists and cardiovascular outcomes in patients with and without prior cardiovascular events: An updated meta-analysis and subgroup analysis of randomized controlled trials. Diabetes Obes Metab. 2019:

[11] 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**: 2895-2909

[12] Birkeland KI, Bodegard J, Norhammar A, *et al.* How representative of a general type 2 diabetes population are patients included in cardiovascular outcome trials with SGLT2 inhibitors? A large European observational study. Diabetes Obes Metab. 2018:

[13] 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**: 1299-1304

[14] Wittbrodt E, Chamberlain D, Arnold SV, Tang F, Kosiborod M. Eligibility of patients with type 2 diabetes for sodium-glucose co-transporter-2 inhibitor cardiovascular outcomes trials: An assessment using the Diabetes Collaborative Registry. Diabetes Obes Metab. 2019; **21**: 1985-1989

[15] Fadini GP, Zatti G, Consoli A, Bonora E, Sesti G, Avogaro A. Rationale and design of the DARWIN-T2D (DApagliflozin Real World evIdeNce in Type 2 Diabetes): A multicenter retrospective nationwide Italian study and crowdsourcing opportunity. Nutr Metab Cardiovasc Dis. 2017; **27**: 1089-1097

[16] Fadini GP, Zatti G, Baldi I, *et al.* Use and effectiveness of dapagliflozin in routine clinical practice: An Italian multicentre retrospective study. Diabetes Obes Metab. 2018; **20**: 1781-1786

[17] Levey AS, Stevens LA, Schmid CH, *et al.* A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009; **150**: 604-612

[18] Ceriello A, Rossi MC, De Cosmo S, *et al.* Overall Quality of Care Predicts the Variability of Key Risk Factors for Complications in Type 2 Diabetes: An Observational, Longitudinal Retrospective Study. Diabetes Care. 2019; **42**: 514-519

[19] Colombo D, Maathuis MH. Order-Independent Constraint-Based Causal Structure Learning. Journal of Machine Learning Research. 2014; **15**: 3921-3962

[20] Broom BM, Do KA, Subramanian D. Model averaging strategies for structure learning in Bayesian networks with limited data. BMC Bioinformatics. 2012; **13 Suppl 13**: S10

[21] Lauritzen SL. Graphical Models. Oxford Statistical Science Series 1996; **Clarendon Press, Oxford.**:

[22] Fadini GP, Frison V, Rigato M, *et al.* Trend 2010-2018 in the clinical use of GLP-1 receptor agonists for the treatment of type 2 diabetes in routine clinical practice: an observational study from Northeast Italy. Acta Diabetol. 2019:

[23] Gerstein HC, McMurray J, Holman RR. Real-world studies no substitute for RCTs in establishing efficacy. Lancet. 2019; **393**: 210-211

[24] Morieri ML, Rigato M, Frison V, *et al.* Fixed versus flexible combination of GLP-1 receptor agonists with basal insulin in type 2 diabetes: A retrospective multicentre comparative effectiveness study. Diabetes Obes Metab. 2019; **21**: 2542-2552

[25] Fadini GP, Sciannameo V, Franzetti I, *et al.* Similar effectiveness of dapagliflozin and GLP-1 receptor agonists concerning combined endpoints in routine clinical practice: A multicentre retrospective study. Diabetes Obes Metab. 2019; **21**: 1886-1894

[26] Mody R, Huang Q, Yu M, *et al.* Adherence, persistence, glycaemic control and costs among patients with type 2 diabetes initiating dulaglutide compared with liraglutide or exenatide once weekly at 12-month follow-up in a real-world setting in the United States. Diabetes Obes Metab. 2018:

[27] Chatterjee S, Davies MJ, Khunti K. What have we learnt from "real world" data, observational studies and meta-analyses. Diabetes Obes Metab. 2018; **20 Suppl 1**: 47-58

[28] Anyanwagu U, Mamza J, Mehta R, Donnelly R, Idris I. Cardiovascular events and all-cause mortality with insulin versus glucagon-like peptide-1 analogue in type 2 diabetes. Heart. 2016; **102**: 1581-1587

[29] Svanstrom H, Ueda P, Melbye M, *et al.* Use of liraglutide and risk of major cardiovascular events: a register-based cohort study in Denmark and Sweden. Lancet Diabetes Endocrinol. 2019; **7**: 106-114

[30] Hong JL, Webster-Clark M, Jonsson Funk M, *et al.* Comparison of Methods to Generalize Randomized Clinical Trial Results Without Individual-Level Data for the Target Population. Am J Epidemiol. 2019; **188**: 426-437

[31] Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of Trial Results Using Inverse Odds of Sampling Weights. Am J Epidemiol. 2017; **186**: 1010-1014

[32] Hong JL, Jonsson Funk M, LoCasale R, *et al.* Generalizing Randomized Clinical Trial Results: Implementation and Challenges Related to Missing Data in the Target Population. Am J Epidemiol. 2018; **187**: 817-827

Table 1. Clinical characteristics of patients treated with GLP-1RA and of those eligible for CVOTs.

	GLP-1RA users in DARWIN-T2D	LEADER	SUSTAIN-6	EXSCEL	REWIND	PIONEER-6	HARMONY
Number	6699	10061	9942	16544	37574	39726	10208
Percentage[a]	5.1	9.4	10.1	13.4	35.8	34.1	9.5
Age, years	61.7±9.5	74.2±8.4	74.2±8.5	70.8±8.6	70.8±7.1	73.7±8.3	73.6±9.1
Sex male, %	54.9	56.5	55.8	67.3	59.0	57.6	68.4
Diabetes duration, years	11.3±7.5	13.6±9.1	13.6±9.1	15.4±9.8	10.7±8.2	14.3±10.0	17.9±10.2
Active smoke, %	19.8	13.5	13.6	17.1	14.3	14.0	16.1
Body mass index, kg/m^2	34.8±6.2	29.1±5.1	29.2±5.2	29.1±4.9	29.8±4.7	29.3±5.2	29.1±4.9
Waist circumference, cm	114.7±13.6	103.9±12.2	104.1±12.4	104.2±11.9	104.5±11.3	104.5±12.6	104.6±11.9
Systolic blood pressure, mm Hg	138.3±18.3	139.1±18.6	139.2±18.7	137.5±18.4	138.4±17.9	138.1±18.7	137.5±18.6
Diastolic blood pressure, mm Hg	80.3±9.9	76.9±9.3	77.0±9.4	76.5±9.2	77.8±9.1	76.3±9.5	75.8±9.2
Heart rate, bpm	78.8±11.9	73.3±11.9	73.5±11.8	71.7±11.6	73.3±11.8	72.8±11.8	70.2±11.0
Fasting plasma glucose, mg/dl	151.2±42.7	151.3±36.9	155.6±42.1	150.2±42.2	136.8±33.7	146.4±47.7	160.4±50.5
HbA1c, %	7.5±1.1	7.8±0.6	8.0±1.0	7.6±0.8	6.9±0.9	7.4±1.3	8.1±1.0
Total cholesterol, mg/dl	169.3±37.5	168.2±38.2	169.0±38.8	160.5±38.3	170.0±37.4	166.8±38.8	158.2±39.6
HDL cholesterol, mg/dl	45.9±12.6	48.7±13.6	48.4±13.5	47.1±13.6	49.8±13.8	49.0±14.5	45.7±13.4
Triglycerides, mg/dl	160.2±87.5	141.5±74.2	144.2±76.5	140.9±85.3	134.6±71.7	137.4±77.6	146.9±88.5
LDL cholesterol, mg/dl	91.9±32.3	91.4±32.2	91.9±32.7	85.3±31.7	93.3±32.1	90.5±32.6	83.1±32.2
eGFR, ml/min/1.73 m^2	87.7±24.1	68.3±21.5	68.5±21.9	73.3±21.6	76.2±19.4	66.7±20.5	68.9±22.1
Albumin excretion rate, mg/g	51.5±147.7	59.5±98.7	61.2±109.5	42.5±108.7	45.3±66.0	57.1±155.0	43.4±132.1
Glucose lowering medications, %							

Insulin	24.9	25.7	27.8	41.0	14.6	43.3	56.4
Metformin	85.9	75.2	73.8	67.8	83.4	62.1	59.5
Sulphonylurea / repaglinide	26.4	52.9	52.0	28.7	32.1	28.0	30.9
Acarbose	2.0	3.5	2.7	1.8	2.2	2.6	2.4
Pioglitazione	9.0	5.8	3.5	3.7	4.0	3.7	3.1
DPP-4 inhibitors	0.2	0.0	0.0	27.6	0.0	0.0	28.0
GLP-1RA	100.0	0.0	0.0	0.0	0.0	0.0	0.0
SGLT-2 inhibitors	0.7	5.7	5.4	4.5	3.4	3.6	5.4
Other therapies, %							
Anti-platelet agents	46.4	58.5	58.5	74.6	51.3	60.9	84.0
Statin	62.9	64.3	63.9	76.1	63.4	64.8	79.4
Renin-angiotensin system blockers	74.5	71.2	71.3	74.0	70.7	72.1	75.9
Calcium channel blockers	25.6	27.0	27.1	28.6	25.7	27.4	29.4
Beta-blockers	31.5	36.4	36.6	44.5	32.7	36.9	49.7
Diuretics	15.8	21.4	21.6	23.7	15.0	25.2	30.5
Complications, %							
Chronic kidney disease	10.0	40.7	40.7	29.1	20.9	44.9	37.4
Albuminuria >30 mg/g	37.3	59.0	59.7	33.5	40.7	57.6	32.1
Retinopathy	15.6	16.1	16.6	24.9	11.4	17.9	31.2
Peripheral neuropathy	14.8	21.2	21.8	25.9	17.2	23.5	30.4
Atherosclerosis obliterans	12.4	27.2	27.7	48.3	13.8	26.1	60.9
Peripheral revascularization	1.2	3.0	3.0	5.2	1.3	2.8	6.4
Diabetic foot	7.6	13.0	13.6	15.9	10.0	12.4	19.3
Stroke / Transient ischemic attack	2.2	9.5	9.8	11.3	4.8	9.3	14.4

Carotid atherosclerosis	39.1	47.5	47.6	51.4	42.1	45.3	54.7
Ischemic heart disease	8.2	20.9	20.9	44.2	11.7	21.0	56.7
Coronary revascularization	6.0	13.5	13.5	29.9	7.5	13.6	37.9
Microangiopathy	43.1	85.4	85.6	61.6	56.5	87.1	69.0
Macroangiopathy	30.4	52.3	52.5	79.8	37.2	50.3	98.2

[a] The percentage refers to the % of patients who were on GLP-1RA in the DAWIN-T2D database (first column) and to the % of patients who would be eligible for each trial in the other columns.

Table 2. Key clinical characteristics of real-world patients compared to CVOT patients. For each CVOT, we show the average clinical characteristics extracted from the respective publications, the characteristics of real-world patients who would be recruited into the CVOT based on inclusion / exclusion (I/E) criteria, and the characteristics of real-world patients sampled for being CVOT-like (Like). For both subgroups of real-world patients, we calculated the absolute standardized mean difference (SMD) as a measure of balance between groups. a SMD≤0.10 is conventionally considered indicative of a good balance.

Variable	**LEADER**	**I/E**	**SMD**	**Like**	**SMD**
Number	9340	10061		1132	
Age, years	64.3 (7.2)	74.2 (8.4)	1.26	64.6 (7.6)	0.05
Sex male, %	64.2	56.5	0.16	65.0	0.02
Diabetes duration	12.8 (8.0)	13.6 (9.1)	0.09	13.5 (8.4)	0.09
HbA1c, %	8.7 (1.5)	7.8 (0.6)	0.80	8.5 (0.8)	0.10
BMI, kg/m^2	32.5 (6.3)	29.1 (5.1)	0.60	32.7 (5.8)	0.03
SBP, mm Hg	135.9 (17.7)	139.1 (18.6)	0.18	137.8 (18.7)	0.10
DBP, mm Hg	77.1 (10.2)	76.9 (9.3)	0.02	78.2 (9.3)	0.10
Heart failure, %	17.9	2.5	0.53	16.0	0.05
Established CVD, %	81.4	55.7	0.58	81.4	0.001
CVD risk factors, %	18.7	28.7	0.24	22.3	0.09
eGFR, ml/min/1.73 m^2	80.4 (21.0)	68.3 (21.5)	0.57	78.0 (26.5)	0.10
Variable	**REWIND**	**I/E**	**SMD**	**Like**	**SMD**
Number	9901	37574		7280	
Age, years	66.2 (6.5)	70.8 (7.1)	0.66	66.7 (6.2)	0.08
Sex male, %	53.9	59.0	0.10	59.3	0.10
Diabetes duration	10.5 (7.2)	10.7 (8.2)	0.02	10.8 (7.1)	0.05
HbA1c, %	7.3 (1.1)	6.9 (0.9)	0.42	7.4 (1.2)	0.06
BMI, kg/m^2	32.3 (5.7)	29.8 (4.7)	0.51	31.9 (5.3)	0.10
SBP, mm Hg	137.0 (17.0)	138.4 (18.0)	0.08	137.2 (15.5)	0.01
DBP, mm Hg	78.0 (9.9)	77.8 (9.1)	0.02	78.7 (8.1)	0.08
Heart failure, %	8.7	1.0	0.36	7.3	0.05
Established CVD, %	31.4	28.2	0.07	30.6	0.02
CVD risk factors, %	68.6	19.9	1.12	63.8	0.10
eGFR, ml/min/1.73 m^2	75.0 (22.1)	75.2 (21.2)	0.009	77.4 (22.2)	0.10
Variable	**SUSTAIN-6**	**I/E**	**SMD**	**Like**	**SMD**
Number	3297	9942		476	
Age, years	64.6 (7.4)	74.2 (8.5)	1.16	65.1 (6.7)	0.07
Sex male, %	60.7	55.8	0.10	64.1	0.07
Diabetes duration	13.9 (8.1)	13.6 (9.1)	0.00	14.5 (6.8)	0.07

HbA1c, %	8.7 (1.5)	8.0 (1.0)	0.61	8.6 (0.7)	0.08
BMI, kg/m^2	32.8 (6.2)	29.2 (5.2)	0.66	32.8 (5.9)	0.004
SBP, mm Hg	135.6 (17.2)	139.2 (18.7)	0.20	136.8 (17.5)	0.07
DBP, mm Hg	77.0 (10.0)	77.0 (9.4)	0.00	77.8 (9.7)	0.08
Heart failure, %	23.6	2.6	0.65	19.5	0.10
Established CVD, %	83.0	55.7	0.62	80.5	0.07
CVD risk factors, %	17.0	29.0	0.29	14.1	0.08
eGFR, ml/min/1.73 m^2	N/A	N/A	N/A	N/A	N/A
Variable	**PIONEER-6**	**I/E**	**SMD**	**Like**	**SMD**
Number	3183	39726		1663	
Age, years	66.0 (7.0)	73.7 (8.3)	0.94	66.6 (7.2)	0.09
Sex male, %	68.4	57.6	0.23	72.9	0.10
Diabetes duration	14.9 (8.5)	14.3 (10.0)	0.06	14.0 (8.6)	0.10
HbA1c, %	8.2 (1.6)	7.4 (1.3)	0.60	8.2 (0.7)	0.01
BMI, kg/m^2	32.3 (6.5)	29.3 (5.2)	0.57	32.0 (3.7)	0.05
SBP, mm Hg	136.0 (18.0)	138.1 (18.7)	0.11	135.7 (14.0)	0.01
DBP, mm Hg	74.0 (21.0)	76.3 (9.5)	0.21	77.0 (8.0)	0.10
Heart failure, %	12.2	2.7	0.37	8.7	0.10
Established CVD, %	84.7	58.1	0.62	80.7	0.10
CVD risk factors, %	15.3	27.8	0.31	19.1	0.10
eGFR, ml/min/1.73 m^2	76.0 (10.0)	66.2 (21.3)	0.47	71.6 (26.0)	0.10
Variable	**EXSCEL**	**I/E**	**SMD**	**Like**	**SMD**
Number	14752	16544		915	
Age, years	62.0 (16.3)	70.8 (8.6)	0.69	62.3 (5.6)	0.02
Sex male, %	62.0	67.3	0.11	62.4	0.008
Diabetes duration	12.0 (7.4)	15.4 (9.8)	0.39	11.5 (6.9)	0.06
HbA1c, %	8.0 (1.2)	7.6 (0.8)	0.40	7.9 (0.7)	0.10
BMI, kg/m^2	31.8 (5.9)	29.1 (4.9)	0.50	31.8 (5.9)	0.006
SBP, mm Hg	N/A	N/A	N/A	N/A	N/A
DBP, mm Hg	N/A	N/A	N/A	N/A	N/A
Heart failure, %	16.2	2.9	0.46	12.6	0.10
Established CVD, %	73.1	64.5	0.19	72.3	0.02
CVD risk factors, %	26.9	27.9	0.02	22.3	0.10
eGFR, ml/min/1.73 m^2	76.3 (22.9)	70.5 (25.5)	0.24	78.4 (31.5)	0.09
Variable	**HARMONY**	**I/E**	**SMD**	**Like**	**SMD**
Number	9463	10208			
Age, years	64.1 (8.7)	73.6 (9.1)	1.07	N/A	N/A
Sex male, %	69.0	68.4	0.01	N/A	N/A

Diabetes duration	14.1 (8.7)	17.9 (10.3)	0.40	N/A	N/A
HbA1c, %	8.7 (1.5)	8.1 (1.0)	0.47	N/A	N/A
BMI, kg/m^2	32.3 (5.9)	29.1 (4.9)	0.59	N/A	N/A
SBP, mm Hg	134.7 (16.5)	137.5 (18.6)	0.16	N/A	N/A
DBP, mm Hg	76.8 (10.1)	75.8 (9.2)	0.10	N/A	N/A
Heart failure, %	20.0	4.3	0.50	N/A	N/A
Established CVD, %	100.0	85.4	0.58	N/A	N/A
CVD risk factors, %	0.0	33.6	1.00	N/A	N/A
eGFR, ml/min/1.73 m^2	79.0 (25.5)	68.9 (22.1)	0.42	N/A	N/A

BMI, body mass index. SBP, systolic blood pressure. DBP, diastolic blood pressure. CVD, cardiovascular disease. eGFR, estimated glomerular filtration rate. N/A, not available. Established CVD and CVD risk factors are defined as described in each trial publication and slightly modified as illustrated in table S1.

Figure 1. Real-world patients and CVOTs. A) For each CVOT, the panels show the absolute standardized mean difference (SMD) between the actual trial population (retrieved from respective publications) and real-world patients selected based on inclusion/exclusion criteria (I/E) or for being CVOT-like (Like). In each plot, a dashed line indicates the SMD threshold of 0.1, indicating good balance. Fractions in brackets refer to the number of key clinical characteristics that are matched between real-world patients selected by I/E and trial characteristics. By design, all characteristics were balanced between CVOT-like patients and the respective CVOT population. B) Proportion of real-wold patients eligible for each CVOT based on I/E or sampled for being CVOT-like.

A