Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, RuiJun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David Madigan, George Hripcsak, Patrick B Ryan

Summary

Background Uncertainty remains about the optimal monotherapy for hypertension, with current guidelines recommending any primary agent among the first-line drug classes thiazide or thiazide-like diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, and non-dihydropyridine calcium channel blockers, in the absence of comorbid indications. Randomised trials have not further refined this choice.

Methods We developed a comprehensive framework for real-world evidence that enables comparative effectiveness and safety evaluation across many drugs and outcomes from observational data encompassing millions of patients, while minimising inherent bias. Using this framework, we did a systematic, large-scale study under a new-user cohort design to estimate the relative risks of three primary (acute myocardial infarction, hospitalisation for heart failure, and stroke) and six secondary effectiveness and 46 safety outcomes comparing all first-line classes across a global network of six administrative claims and three electronic health record databases. The framework addressed residual confounding, publication bias, and p-hacking using large-scale propensity adjustment, a large set of control outcomes, and full disclosure of hypotheses tested.

Findings Using 4·9 million patients, we generated 22 000 calibrated, propensity-score-adjusted hazard ratios (HRs) comparing all classes and outcomes across databases. Most estimates revealed no effectiveness differences between classes; however, thiazide or thiazide-like diuretics showed better primary effectiveness than angiotensin-converting enzyme inhibitors: acute myocardial infarction (HR 0·84, 95% CI 0·75–0·95), hospitalisation for heart failure (0·83, 0·74–0·95), and stroke (0·83, 0·74–0·95) risk while on initial treatment. Safety profiles also favoured thiazide or thiazide-like diuretics over angiotensin-converting enzyme inhibitors. The non-dihydropyridine calcium channel blockers were significantly inferior to the other four classes.

Interpretation This comprehensive framework introduces a new way of doing observational health-care science at scale. The approach supports equivalence between drug classes for initiating monotherapy for hypertension—in keeping with current guidelines, with the exception of thiazide or thiazide-like diuretics superiority to angiotensin-converting enzyme inhibitors and the inferiority of non-dihydropyridine calcium channel blockers.

Funding US National Science Foundation, US National Institutes of Health, Janssen Research & Development, IQVIA, South Korean Ministry of Health & Welfare, Australian National Health and Medical Research Council.

Introduction

Patients and physicians have a wide range of pharmacological options to treat hypertension but little guidance on which specific first-line agent to initiate. The 2017 American College of Cardiology/American Heart Association (ACC/AHA) Blood Pressure Treatment Guidelines[1] endorse any thiazide or thiazide-like diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, or calcium channel blockers unless contraindicated. Similar non-specificity emerges from the 2018 European Society of Cardiology/European Society of Hypertension (ESC/ESH) Guidelines, with the further inclusion of β blockers.[2]

These recommendations derive largely from earlier randomised clinical trials (RCTs) that provided direct comparisons between a few agents, not drug classes, and often did not restrict to therapy initiation. For example, the largest head-to-head RCT of antihypertensives, the ALLHAT trial,[3] enrolled patients from February, 1994, to January, 1998, more than two decades ago, evaluated three representative agents and a majority of participants had been previously treated. Moreover, most studies considered in the 2017 ACC/AHA Guidelines systematic review[4] were done before 2000.

The 2017 Cochrane Review[5] of first-line therapy for hypertension, an update from 2009, found no new RCTs to include. Their review concludes that "first-line low-dose

Published **Online** October 24, 2019 https://doi.org/10.1016/S0140-6736(19)32317-7

See **Online/Comment** https://doi.org/10.1016/S0140-6736(19)32461-4

Department of Biostatistics, Fielding School of Public Health (Prof M A Suchard MD, M J Schuemie PhD), **and Department of Biomathematics, David Geffen School of Medicine at UCLA** (Prof M A Suchard), **University of California, Los Angeles, CA, USA; Epidemiology Analytics, Janssen Research & Development, Titusville, NJ, USA** (M J Schuemie, P B Ryan PhD); **Department of Medicine, Yale University School of Medicine, New Haven, CA, USA** (Prof H M Krumholz MD); **Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea** (S C You MD); **Department of Medicine, Weill Cornell Medical College, New York, NY, USA** (R Chen MD); **Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA** (R Chen, Prof G Hripcsak MD, P B Ryan); **Quality Use of Medicines and Pharmacy Research Centre, Sansom Institute, University of South Australia, Adelaide, SA, Australia** (N Pratt PhD); **Real World Analytics Solutions, IQVIA, Durham, NC, USA** (C G Reich MD); **Georgia Tech Research Institute, Georgia Tech College of Computing, Atlanta, GA, USA** (J Duke MD); **Department of Statistics, Columbia University, New York, NY, USA** (Prof D Madigan PhD); **and Medical Informatics Services, New York-Presbyterian Hospital, New York, NY, USA** (Prof G Hripcsak)

Correspondence to:
Prof Marc A Suchard,
Department of Biostatistics,
Fielding School of Public Health,
University of California,
Los Angeles, CA 90095, USA
msuchard@ucla.edu

Research in context

Evidence before this study

2017 American College of Cardiology/American Heart Association Blood Pressure Treatment Guidelines recommend initiating monotherapy for hypertension with any primary agent among five first-line drug classes based on a systematic review of randomised trials. Similar non-specificity emerges from the 2018 European Society of Cardiology/European Society of Hypertension Guidelines. The largest such trial, ALLHAT, enrolled patients more than two decades ago, only evaluated three representative agents, and a majority of participants had been previously treated for hypertension. We lack contemporary knowledge of the real-world comparative effectiveness of common antihypertensive drugs with respect to outcomes and the safety trade-offs among these class options for treatment initiation.

Added value of this study

Our study uses state-of-the-art methods to control for residual confounding, publication bias, and p-hacking in real-world evidence studies, and shows generally comparable effectiveness between drug classes across nine international health databases. However, effectiveness and safety benefits suggest initiating with a thiazide or thiazide-like diuretic over an angiotensin converting-enzyme inhibitor, the most common initiating monotherapy across databases. Non-dihydropyridine calcium channel blockers are also inferior to thiazide or thiazide-like diuretics, angiotensin converting-enzyme inhibitors, angiotensin receptor blockers, and dihydropyridine calcium channel blockers.

Implications of all the available evidence

Initiating with a thiazide instead of an angiotensin converting-enzyme inhibitor carries potential to avoid many major cardiovascular events and warrants further study.

thiazides reduced all morbidity and mortality outcomes in adult patients with moderate to severe primary hypertension. First-line angiotensin-converting enzyme inhibitors and calcium channel blockers may be similarly effective, but the evidence was of lower quality". Thus, there remains uncertainty and, unfortunately, we do not have contemporary knowledge of the real-world comparative effectiveness of common antihypertensive drugs with respect to outcomes—and the safety trade-offs among these options.

Accordingly, we developed the open-science, large-scale evidence generation and evaluation across a network of databases for hypertension (LEGEND-HTN) study to compare common antihypertensive drug treatments by a systematic, large-scale analysis across nine observational databases from the Observational Health Data Science and Informatics (OHDSI) distributed data network.[6] This novel approach used massive data across several countries and synthesised tens of thousands of comparisons with analytic techniques to minimise residual confounding. In contrast to a single comparison approach, LEGEND-HTN provides a comprehensive view of the findings and their consistency across populations, drugs, and outcomes, and by design avoids the harms of publication bias or overemphasising a single observational analysis subject to p-hacking. We report results comparing monotherapy drug classes from participating data sources through November, 2018, covering patients from July, 1996, to March, 2018.

For more on the **Observational Medical Outcomes Partnership model** see https://github.com/OHDSI/CommonDataModel

See **Online** for appendix

Methods

Study design

LEGEND-HTN systematically executed a large-scale comparative effectiveness and safety study across six administrative claims and three electronic health record databases standardised to OHDSI's Observational Medical Outcomes Partnership common data model version 5 that maps international coding systems into standard vocabulary concepts. The claims databases were IBM MarketScan Commercial Claims and Encounters (US employer-based private payer; patient aged 65 years or older), Optum ClinFormatics (US private payer; primarily aged 65 years or younger), IBM MarketScan Medicare Supplemental Beneficiaries (US retirees; patients aged >65 years), IBM MarketScan Multi-state Medicaid (US Medicaid enrolees; all ages), Japan Medical Data Center (Japan private payer; patients aged 18–65 years), and South Korea National Health Insurance Service/National Sample Cohort (South Korea; all ages). The electronic health records are Optum Pan-Therapeutic (US health systems; all ages), IMS/IQVIA Disease Analyzer Germany (German ambulatory care; all ages), and Columbia University Medical Center (US academic health system; all ages). Database details are included in the appendix (pp 2–4). All data partners had previous institutional review board approval or exemption for their participation.

Within each database source, we used a retrospective, comparative new-user cohort design.[7,8] We considered patients new users if their first observed treatment for hypertension was monotherapy with any active ingredient within the five drug classes listed as primary agents in the 2017 AHA/ACC Guidelines[1] (ie, thiazide or thiazide-like diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, or non-dihydropyridine calcium channel blockers). We required patients to have at least 1 year of previous database observation before first exposure and a recorded hypertension diagnosis at or within the year preceding treatment initiation.

Outcomes

We studied 55 outcomes of interest, including both effectiveness and safety endpoints. We divided effectiveness outcomes into three primary endpoints (ie, acute myocardial infarction, hospitalisation for heart failure, and stroke, on the basis of their use in the 2017 AHA/ACC Guidelines systematic review),[4] and six secondary effectiveness outcomes that major hypertension treatment RCTs have considered.[3,9,10] The 46 safety outcomes were antihypertensive drug side-effects, including angioedema, cough, electrolyte imbalance, gout, diarrhoea, and kidney disease. We constructed all outcomes on the basis of previously published phenotypes (appendix pp 5–9), and each typically involved one or more diagnosis codes in the inpatient or outpatient setting. Full and reproducible cohort instantiation details for myocardial infarction, hospitalisation for heart failure, and stroke in any Observational Medical Outcomes Partnership database and links to computer-readable details for the remaining outcomes are provided in the appendix (pp 9–13).

For each outcome, we excluded patients with events before initiation, and defined patient time-at-risk as either on-treatment analysis, which follows patients from 1 day after treatment initiation until they first discontinue their initial therapy choice or their record ends, or intention-to-treat analysis, which follows patients until their record ends. We constructed these continuous drug exposures from the available longitudinal data by grouping sequential prescriptions that had a gap of less than 30 days between them. We present further details on exposure and outcome cohort construction and standardised execution across the network in the appendix (pp 4–9).

Statistical analysis

To adjust for potential measured confounding and improve the balance between drug class cohorts, we built propensity score models[11] for each class pair and data source using a consistent data-driven process through regularised regression.[12] This process allowed the data to decide which combinations of a large set of predefined baseline patient characteristics, including demographics and previous conditions, drug exposures, procedures, and health-service-use behaviours were most predictive of treatment assignment (appendix p 27). The number of potential characteristics differed across class pair and data source, ranging from 7515 (angiotensin receptor blocker *vs* dihydropyridine calcium channel blockers in Japan Medical Data Center) to 70 784 (angiotensin-converting enzyme inhibitors *vs* dihydropyridine calcium channel blockers in Optum ClinFormatics). We stratified or variable-ratio matched patients by propensity score and used Cox proportional hazards models to estimate hazard ratios (HRs) between alternative target and comparator treatments for the risk of each outcome in each data source. We aggregated HR estimates across data sources to produce

	Patients	On-treatment time (days)	Total follow-up time (days)
THZ	830 608	..	..
CCAE	305 741	95 (29–486)	733 (300–2247)
Optum	201 325	90 (30–478)	740 (303–2183)
MDCR	36 683	116 (29–612)	1025 (418–2779)
MDCD	34 743	59 (30–245)	553 (253–1828)
NHIS/NSC	6454	29 (6–414)	2555 (1397–3680)
PanTher	234 274	89 (89–198)	1245 (547–2534)
IMSG	5113	100 (50–287)	1310 (528–2909)
CUMC	6275	250 (51–1537)	1807 (752–3652)
ACEi	2 373 007	..	..
CCAE	779 041	116 (38–530)	675 (282–1960)
Optum	563 419	118 (30–555)	722 (298–2130)
MDCR	101 610	152 (58–652)	831 (365–2423)
MDCD	66 185	78 (30–329)	578 (262–1879)
NHIS/NSC	5317	67 (27–525)	2756 (1733–3738)
PanTher	737 065	89 (89–200)	1099 (459–2313)
IMSG	109 799	100 (50–402)	1196 (508–2627)
CUMC	10 571	104 (30–1225)	1388 (511–3390)
ARB	752 492	..	..
CCAE	230 002	147 (54–628)	699 (288–2149)
Optum	170 852	146 (46–640)	694 (292–2100)
MDCR	31 647	195 (83–779)	953 (401–2661)
MDCD	7764	87 (30–347)	548 (249–2008)
JMDC	53 532	218 (58–983)	793 (354–1865)
NHIS/NSC	16 286	128 (29–1004)	1475 (706–3010)
PanTher	207 097	89 (37–187)	1017 (395–2259)
IMSG	29 951	98 (56–427)	974 (414–2323)
CUMC	5361	90 (30–500)	1153 (482–2673)
dCCB	798 540	..	..
CCAE	217 684	89 (29–456)	613 (254–1803)
Optum	169 209	91 (30–515)	660 (272–2014)
MDCR	38 514	143 (47–654)	768 (341–2240)
MDCD	34 860	53 (30–238)	494 (217–1548)
JMDC	51 770	136 (30–741)	649 (291–1665)
NHIS/NSC	33 050	60 (14–815)	2101 (1124–3422)
PanTher	227 899	89 (84–187)	919 (336–2139)
IMSG	18 262	100 (50–328)	1176 (471–2632)
CUMC	7292	90 (30–768)	1099 (329–2971)
ndCCB	138 944	..	..
CCAE	33 382	93 (29–528)	719 (298–2265)
Optum	38 831	119 (30–663)	780 (307–2249)
MDCR	10 613	134 (36–676)	819 (352–2513)
MDCD	4248	61 (30–303)	657 (275–2190)
PanTher	51 870	89 (29–163)	1272 (552–2527)

Data are n or median (IQR). When executing comparative studies, we excluded database populations with fewer than 2500 new users. THZ=thiazide or thiazide-like diuretics. CCAE=IBM MarketScan Commercial Claims and Encounters. Optum=Optum ClinFormatics. MDCR=IBM MarketScan Medicare Supplemental Beneficiaries. MDCD=IBM MarketScan Multi-state Medicaid. NHIS/NSC=South Korea National Health Insurance Service/National Sample Cohort. PanTher=Optum Pan-Therapeutic. IMSG=IMS/IQVIA Disease Analyzer Germany. CUMC=Columbia University Medical Center. ACEi=angiotensin-converting enzyme inhibitors. ARB=angiotensin receptor blockers. JMDC=Japan Medical Data Center. dCCB=dihydropyridine calcium channel blockers. ndCCB=non-dihydropyridine calcium channel blockers.

Table 1: **Population size and follow-up time for each first-line blood pressure lowering drug class within each database**

	Before stratification			After stratification		
	THZ	ACEi	Standardised difference	THZ	ACEi	Standardised difference
Age group (years)						
10–14	0·1%	0·2%	–0·02	0·1%	0·1%	0
15–19	0·6%	0·7%	–0·02	0·7%	0·7%	0
20–24	1·6%	1·4%	0·02	1·5%	1·4%	0·01
25–29	3·5%	2·6%	0·06	2·7%	2·8%	0
30–34	6·6%	5·0%	0·07	5·4%	5·4%	0
35–39	9·8%	8·1%	0·06	8·5%	8·5%	0
40–44	13·4%	12·1%	0·04	12·3%	12·4%	0
45–49	16·3%	16·1%	0·01	15·9%	16·2%	–0·01
50–54	17·7%	18·7%	–0·03	18·4%	18·5%	0
55–59	16·2%	18·3%	–0·06	18·0%	17·8%	0
60–64	13·2%	15·5%	–0·06	15·3%	15·0%	0·01
65–69	1·1%	1·3%	–0·02	1·3%	1·3%	0
Sex						
Female	60·7%	38·4%	0·46	45·2%	44·7%	0·01
Male	39·3%	61·6%	–0·46	54·8%	55·3%	–0·01
Medical history (general)						
Acute respiratory disease	26·1%	24·5%	0·04	25·5%	25·0%	0·01
Attention deficit hyperactivity disorder	1·1%	1·2%	–0·01	1·2%	1·1%	0
Chronic liver disease	1·1%	1·5%	–0·03	1·3%	1·4%	0
Chronic obstructive lung disease	1·4%	1·8%	–0·03	1·7%	1·7%	0
Crohn's disease	0·3%	0·3%	0	0·3%	0·3%	0
Dementia	0·1%	0·1%	0	0·2%	0·1%	0·01
Depressive disorder	8·1%	7·4%	0·03	7·9%	7·6%	0·01
Diabetes	4·6%	18·3%	–0·44	13·5%	14·5%	–0·03
Gastro-oesophageal reflux disease	7·5%	7·8%	–0·01	7·8%	7·8%	0
Gastrointestinal haemorrhage	1·6%	1·7%	–0·01	1·8%	1·7%	0·01
HIV infection	0·3%	0·2%	0·01	0·2%	0·2%	0
Hyperlipidaemia	25·5%	36·1%	–0·23	33·0%	33·2%	0
Lesion of liver	0·2%	0·2%	–0·01	0·3%	0·2%	0·01
Obesity	10·0%	8·5%	0·05	9·1%	8·8%	0·01
Osteoarthritis	10·7%	11·3%	–0·02	11·5%	11·2%	0·01
Pneumonia	1·4%	1·5%	0	1·6%	1·5%	0·01
Psoriasis	0·9%	1·0%	–0·02	1·0%	1·0%	0
Renal impairment	0·5%	1·1%	–0·06	1·0%	0·9%	0·01
Rheumatoid arthritis	0·8%	0·8%	0·01	0·8%	0·8%	0
Schizophrenia	0·1%	0·1%	0	0·1%	0·1%	0
Ulcerative colitis	0·2%	0·3%	–0·01	0·3%	0·3%	0
Urinary tract infectious disease	6·4%	5·1%	0·05	5·7%	5·5%	0·01
Viral hepatitis C	0·3%	0·4%	–0·01	0·4%	0·4%	0
Visual system disorder	14·9%	15·5%	–0·02	15·5%	15·5%	0
Medical history (cardiovascular disease)						
Atrial fibrillation	0·3%	0·4%	–0·03	0·4%	0·4%	0
Cerebrovascular disease	1·0%	1·7%	–0·06	1·6%	1·6%	0
Coronary arteriosclerosis	1·0%	2·2%	–0·10	1·9%	1·9%	0
Heart disease	6·5%	9·0%	–0·09	8·7%	8·4%	0·01

(Table 2 continues on next page)

meta-analytic estimates using a random-effects meta-analysis.[13] For the monotherapy initiation of the five drug classes (ten pairwise comparisons) to study 55 outcomes in nine databases (plus one meta-analysis) using two time-at-risk definitions and two propensity score-adjustment approaches, we generated 22 000 effect estimates.

Residual study bias from unmeasured and systematic sources can still exist in observational studies after controlling for measured confounding.[14,15] Therefore, for each effect estimate, we did negative control outcome experiments, where the null hypothesis of no effect was believed to be true using 76 controls (appendix pp 13–15) identified through a data-rich algorithm.[16] We used the empirical null distributions and synthetic positive controls[17] to calibrate each HR estimate, its 95% CI, and the p value to reject the null hypothesis of no differential effect. A HR was significantly different from the null value when its calibrated 95% CI did not include this value. This corresponds to a calibrated p of less than 0·05 without correcting for multiple testing.

Finally, for each of the 22 000 target–comparator–outcome–database–analysis combinations, we report full study diagnostics and results. These include power calculations estimating minimum detectable relative risk, preference score (a transformation of propensity score that adjusts for prevalence differences between populations) distributions to evaluate empirical equipoise[18] and population generalisability, patient characteristics to evaluate cohort balance before and after propensity score adjustment, negative and positive control calibration plots to assess residual bias, and Kaplan-Meier plots to examine HR proportionality assumptions. We defined target and comparator cohorts to stand in empirical equipoise if the majority of patients in both carry preference scores between 0·3 and 0·7 and to achieve sufficient balance if all after-adjustment baseline characteristics returned absolute standardised mean differences of less than 0·1.

Because of the potential confounding effect of blood pressure, and to better understand the effect of the lack of baseline blood pressure measurements on effectiveness and safety estimation that arises in administrative claims and some electronic health record data, we did a non-prespecified sensitivity analysis within the Optum Pan-Therapeutic database. This electronic health record records systolic and diastolic blood pressure for most participants. For each class pair, we first rebuilt propensity score models where we additionally included baseline blood pressure measurements as patient characteristics, stratified or matched patients under the new propensity score models that directly adjusted for potential blood pressure confounding, and then estimated effectiveness and safety HRs.

We did this study using the open-source OHDSI CohortMethod R package with large-scale analytics achieved through the Cyclops R package.[19] The pre-specified LEGEND-HTN protocol and end-to-end open

and executable source code are available online. We developed an interactive LEGEND website to promote transparency and allow for sharing and exploration of the complete result set online. For clarity, we present here principal comparisons and outcomes under an on-treatment, propensity score-stratified design. All comparisons, outcomes, databases, and analysis choices of interest are presented on the website and in the appendix (pp 31–38).

Role of the funding source

None of the funding sources (Janssen Research & Development, IQVIA, US National Science Foundation, US National Institutes of Health, South Korean Ministry of Health & Welfare, and Australian National Health and Medical Research Council) had input in the design, execution, interpretation of results or decision to publish.

Results

LEGEND-HTN included longitudinal claims and electronic health record data from 4 893 591 patients, 48% of whom initiated an angiotensin-converting enzyme inhibitor, 17% a thiazide or thiazide-like diuretic, 16% a dihydropyridine calcium channel blocker, 15% an angiotensin receptor blocker, and 3% a non-dihydropyridine calcium channel blocker (table 1). The IBM MarketScan Commercial Claims and Encounters, Optum Pan-Therapeutic, and Optum ClinFormatics databases contributed the most patients to the study across all five drug classes. Median on-treatment time-at-risk for patients varied by drug class and database between 1 and 7 months, but in most databases 25% of the patients were exposed to their first drug class for more than 1 year. Median overall follow-up for patients was more than 2 years for most databases, with 25% of patients having more than 5 years of follow-up in each drug class. Individual drug ingredients within each class are provided in the appendix (p 5). The majority of angiotensin-converting enzyme inhibitors new users started on lisinopril (80%), thiazide or thiazide-like diuretic new users on hydrochlorothiazide (94%), angiotensin receptor blocker new users on losartan (45%), dihydropyridine calcium channel blockers new users on amlodipine (85%), and non-dihydropyridine calcium channel blockers new users on diltiazem (62%).

Patient baseline characteristics for one target–comparator–database combination, comparing patients initiating thiazide or thiazide-like diuretic (target) with patients initiating angiotensin-converting enzyme inhibitors (comparator) in the IBM MarketScan Commercial Claims and Encounters database, are presented in table 2. Before propensity score stratification, angiotensin-converting enzyme inhibitors new users were more likely to be male, have diabetes, hyperlipidaemia, arteriosclerosis, or heart disease relative to patients initiating a thiazide or thiazide-like diuretic. After stratification, the thiazide or thiazide-like diuretic and angiotensin-converting enzyme inhibitors populations were well balanced on all 56 535 baseline patient characteristics. Patient baseline characteristics for the remaining pairwise class comparisons in IBM MarketScan

	Before stratification			After stratification		
	THZ	ACEi	Standardised difference	THZ	ACEi	Standardised difference
(Continued from previous page)						
Heart failure	0·3%	0·5%	−0·02	0·5%	0·4%	0
Ischaemic heart disease	0·9%	1·7%	−0·07	1·4%	1·5%	−0·01
Peripheral vascular disease	3·3%	4·1%	−0·04	4·1%	3·9%	0·01
Pulmonary embolism	0·2%	0·2%	0	0·2%	0·2%	0·01
Venous thrombosis	1·0%	1·0%	0	1·0%	1·0%	0
Medical history (neoplasms)						
Haematological neoplasm	0·4%	0·5%	−0·01	0·6%	0·5%	0·01
Malignant lymphoma	0·2%	0·2%	−0·01	0·2%	0·2%	0·01
Malignant neoplasm of anorectum	0·1%	0·1%	−0·01	0·1%	0·1%	0
Malignant neoplastic disease	3·8%	4·2%	−0·02	4·4%	4·1%	0·01
Malignant tumour of breast	1·0%	0·7%	0·03	0·9%	0·8%	0·01
Malignant tumour of colon	0·2%	0·2%	−0·01	0·2%	0·2%	0
Malignant tumour of lung	0·1%	0·1%	0	0·1%	0·1%	0
Malignant tumour of urinary bladder	0·1%	0·1%	−0·01	0·1%	0·1%	0
Primary malignant neoplasm of prostate	0·3%	0·5%	−0·03	0·5%	0·5%	0
Medication use						
Antibacterials for systemic use	50·7%	48·8%	0·04	50·1%	49·3%	0·02
Antidepressants	19·1%	17·7%	0·04	18·6%	18·2%	0·01
Antiepileptics	6·0%	6·2%	−0·01	6·3%	6·2%	0
Anti-inflammatory and antirheumatic products	26·3%	24·0%	0·05	25·1%	24·6%	0·01
Antineoplastic agents	1·5%	1·4%	0	1·5%	1·5%	0·01
Antipsoriatics	0·4%	0·4%	0	0·4%	0·4%	0
Antithrombotic agents	2·2%	3·3%	−0·06	3·2%	3·0%	0·01
β blockers	0·4%	0·5%	−0·01	0·5%	0·5%	0
Calcium channel blockers	0	0	0	0	0	−0·01
Diuretics	0	0	0	0	0	0
Drugs for acid-related disorders	14·0%	14·1%	0	14·4%	14·1%	0·01
Drugs for obstructive airway diseases	20·3%	18·1%	0·06	19·0%	18·8%	0·01
Drugs used in diabetes	3·1%	15·6%	−0·44	10·9%	12·1%	−0·04
Immunosuppressants	1·5%	1·5%	0	1·5%	1·5%	0
Lipid-modifying agents	13·6%	24·6%	−0·28	21·0%	21·6%	−0·02
Opioids	16·0%	15·2%	0·02	15·9%	15·5%	0·01
Psycholeptics	18·2%	17·6%	0·02	18·4%	17·8%	0·02
Psychostimulants, agents used for attention deficit hyperactivity disorder, and nootropics	3·1%	2·9%	0·02	3·1%	2·9%	0·01

Less extreme standardised difference of population proportions through stratification suggest improved balance between patient cohorts through propensity score adjustment. THZ=thiazide or thiazide-like diuretics. ACEi=angiotensin-converting enzyme inhibitors.

Table 2: **Baseline patient characteristics for new users of THZ and ACEi in the IBM MarketScan Commercial Claims and Encounters database**

For the **OHDSI CohortMethod R package** see https://github.com/OHDSI/CohortMethod

	Comparator	Acute myocardial infarction	Hospitalisation for heart failure	Stroke
THZ	ACEi	0·84 (0·75–0·95), 0·01	0·83 (0·74–0·95), 0·01	0·83 (0·74–0·95), 0·01
THZ	ARB	0·93 (0·81–1·11), 0·41	0·90 (0·79–1·06), 0·19	0·93 (0·80–1·11), 0·41
THZ	dCCB	0·90 (0·81–1·02), 0·14	0·90 (0·80–1·04), 0·18	0·89 (0·79–1·03), 0·14
THZ	ndCCB	0·70 (0·59–0·84), <0·01	0·58 (0·52–0·65), <0·01	0·78 (0·71–0·87) 0·01
ACEi	ARB	1·11 (0·95–1·32), 0·20	1·05 (0·88–1·26), 0·60	1·07 (0·92–1·27), 0·38
ACEi	dCCB	1·08 (0·96–1·22), 0·18	1·08 (0·94–1·25), 0·24	1·05 (0·93–1·21), 0·38
ACEi	ndCCB	0·87 (0·77–1·00), 0·04	0·68 (0·60–0·78), <0·01	0·89 (0·82–0·98), 0·02
ARB	dCCB	0·95 (0·80–1·14), 0·69	1·04 (0·86–1·26), 0·66	0·99 (0·83–1·19), 0·93
ARB	ndCCB	0·78 (0·69–0·91), 0·01	0·71 (0·64–0·80), <0·01	0·84 (0·73–0·97), 0·05
dCCB	ndCCB	0·84 (0·76–0·93), <0·01	0·73 (0·68–0·78), <0·01	0·87 (0·79–0·96), 0·01

Data are HR (95% CI), p value. Estimates were calibrated to reduce residual bias and report the HR for patients in the target cohort relative to comparator cohort. HRs of less than 1 favour target. THZ=thiazide or thiazide-like diuretics. ACEi=angiotensin-converting enzyme inhibitors. ARB=angiotensin receptor blockers. dCCB=dihydropyridine calcium channel blockers. ndCCB=non-dihydropyridine calcium channel blockers. HR=hazard ratio.

Table 3: **Meta-analytic HR estimates and their 95% CIs comparing the relative risk of highlighted cardiovascular efficacy outcomes between target and comparator in new users of first-line antihypertensive drug classes**

For the **protocol** see https://github.com/OHDSI/Legend/blob/master/Documents/OHDSI%20Legend%20Protocol%20Hypertension%20V03.docx

For the **LEGEND-HTN source code** see https://github.com/OHDSI/Legend

For the **LEGEND website** see http://data.ohdsi.org/LegendBasicViewer

Commercial Claims and Encounters are provided in the appendix (pp 18–26). We found non-dihydropyridine calcium channel blockers new users to have a higher baseline prevalence of atrial fibrillation and other heart diseases than other class users, whereas women new users of dihydropyridine calcium channel blockers were more likely to be pregnant than angiotensin-converting enzyme inhibitors or angiotensin receptor blocker (classes for which use during pregnancy is specifically contraindicated) new users (appendix pp 29–30). Histograms displaying baseline systolic and diastolic blood pressure for new users across all drug classes in the Optum Pan-Therapeutic database are given in the appendix (p 81). Thiazide or thiazide-like diuretic new users had the highest median blood pressure of 142/88 mm Hg (IQR 130/80–152/95), followed by dihydropyridine calcium channel blockers (141/84 mm Hg, 130/76–155/94), angiotensin-converting enzyme inhibitors (140/84 mm Hg, 128/76–152/92), angiotensin receptor blockers (138/82 mm Hg, 126/74–150/90), and non-dihydropyridine calcium channel blockers (133/80 mm Hg, 122/70–146/87).

For five data sources (IBM MarketScan Commercial Claims and Encounters, IBM MarketScan Medicare Supplemental Beneficiaries, IMS/IQVIA Disease Analyzer Germany, Japan Medical Data Center, Columbia University Medical Center), all executed class comparisons were in empirical equipoise (preference score distributions in IBM MarketScan Commercial Claims and Encounters are given in the appendix, p 28). IBM MarketScan Multi-state Medicaid, Optum ClinFormatics, Optum Pan-Therapeutic, and South Korea National Health Insurance Service showed less equipoise for comparisons involving angiotensin receptor blockers or non-dihydropyridine calcium channel blockers. However, in general, propensity score adjustment achieved sufficient covariate balance to reduce concerns that measured estimated effects of baseline-confounding biases (appendix pp 29–30). Finally, before calibration, nominal 95% CIs covered 86·7% of control estimates across all comparisons; after calibration, they covered 96·7%.

The meta-analytic comparative effect estimates for our primary effectiveness outcomes—acute myocardial infarction, hospitalisation for heart failure, and stroke—are presented in table 3. More than half of the comparisons showed no significant difference between classes at a nominal 5% type I error rate. However, thiazide or thiazide-like diuretics showed a significantly lower risk of all three outcomes relative to angiotensin-converting enzyme inhibitors (acute myocardial infarction HR 0·84, 95% CI 0·75–0·95; heart failure 0·83, 0·74–0·95; and stroke 0·83, 0·74–0·95) with an approximate 15% lower event rate. Patient counts, observation time, and events for pairwise class comparisons under the primary effectiveness outcomes are given in the appendix (pp 31–36).

Thiazide or thiazide-like diuretics also showed a significantly lower risk of acute myocardial infarction, hospitalisation for heart failure, and stroke relative to non-dihydropyridine calcium channel blockers (table 3). We observed no significant differences in these outcomes between thiazide or thiazide-like diuretics and either angiotensin receptor blockers or dihydropyridine calcium channel blockers; however, we found that the two subtypes of calcium channel blockers exhibited significantly differential hazards, with dihydropyridine calcium channel blockers having a lower risk of acute myocardial infarction, hospitalisation for heart failure, and stroke relative to non-dihydropyridine calcium channel blockers. Finally, we observed no differences in these three primary effectiveness outcomes between angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, and dihydropyridine calcium channel blockers.

The meta-analytic comparative effect estimates across all nine effectiveness outcomes are presented in figure 1. Seven of these outcomes showed a significantly decreased HR in favour of thiazide or thiazide-like diuretics as compared with angiotensin-converting enzyme inhibitors. We observed no significant differences in outcomes in the remaining comparisons, with the marked exception of non-dihydropyridine calcium channel blockers, which underperformed all other drug classes. Meta-analytic estimates are further stratified into their individual data source-specific contributions for one exemplar outcome: major cardiovascular events (a composite outcome based on ALLHAT[3] of acute myocardial infarction, hospitalisation for heart failure, stroke, and sudden cardiac death; figure 1). In all cardiovascular event comparisons, data sources had relatively consistent estimates, with an I^2 of less than 40% indicating low heterogeneity. In comparing thiazide or thiazide-like diuretics and angiotensin-converting enzyme inhibitors, we observed that three databases independently had significantly decreased effect

estimates, and the meta-analysis allowed greater precision around the estimate (HR 0·84, 95% CI 0·75–0·95) than any one source alone. Relative to non-dihydropyridine calcium channel blockers, we again saw that thiazide or thiazide-like diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, and dihydropyridine calcium channel blockers all showed decreased risks of cardiovascular events, with two or more sources contributing significant effect estimates to the meta-analysis.

Meta-analytic effect estimates for all 46 safety outcomes comparing thiazide or thiazide-like diuretics with angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, and non-dihydropyridine calcium channel blockers are presented in figure 2. The remaining comparisons are in the appendix (pp 69–71). Relative to other drug classes, thiazide or thiazide-like diuretics had a significantly higher risk of hypokalaemia: HR 2·8, 95% CI 2·2–3·6 versus angiotensin-converting enzyme inhibitors; 2·9, 2·2–4·3 versus angiotensin receptor blockers; 1·9, 1·6–2·4 versus dihydropyridine calcium channel blockers; and 1·8, 1·5–2·1 versus non-dihydropyridine calcium channel blockers; and, correspondingly, a significantly lower risk of hyperkalaemia. Thiazide or thiazide-like diuretics also showed a significantly higher risk of hyponatraemia compared with other drug classes. As expected, the risk of angio-oedema and cough was significantly increased for angiotensin-converting enzyme inhibitors new users. The resulting propensity score-adjusted, calibrated HR for angio-oedema for thiazide or thiazide-like diuretic versus angiotensin-converting enzyme inhibitors new users was 0·44 (95% CI 0·35–0·57). Across all disease categories, 16 further safety outcomes occurred at a significantly higher rate in angiotensin-converting enzyme inhibitors, as compared with thiazide or thiazide-like diuretic new users including mortality, gastrointestinal side-effects, and renal disorders.

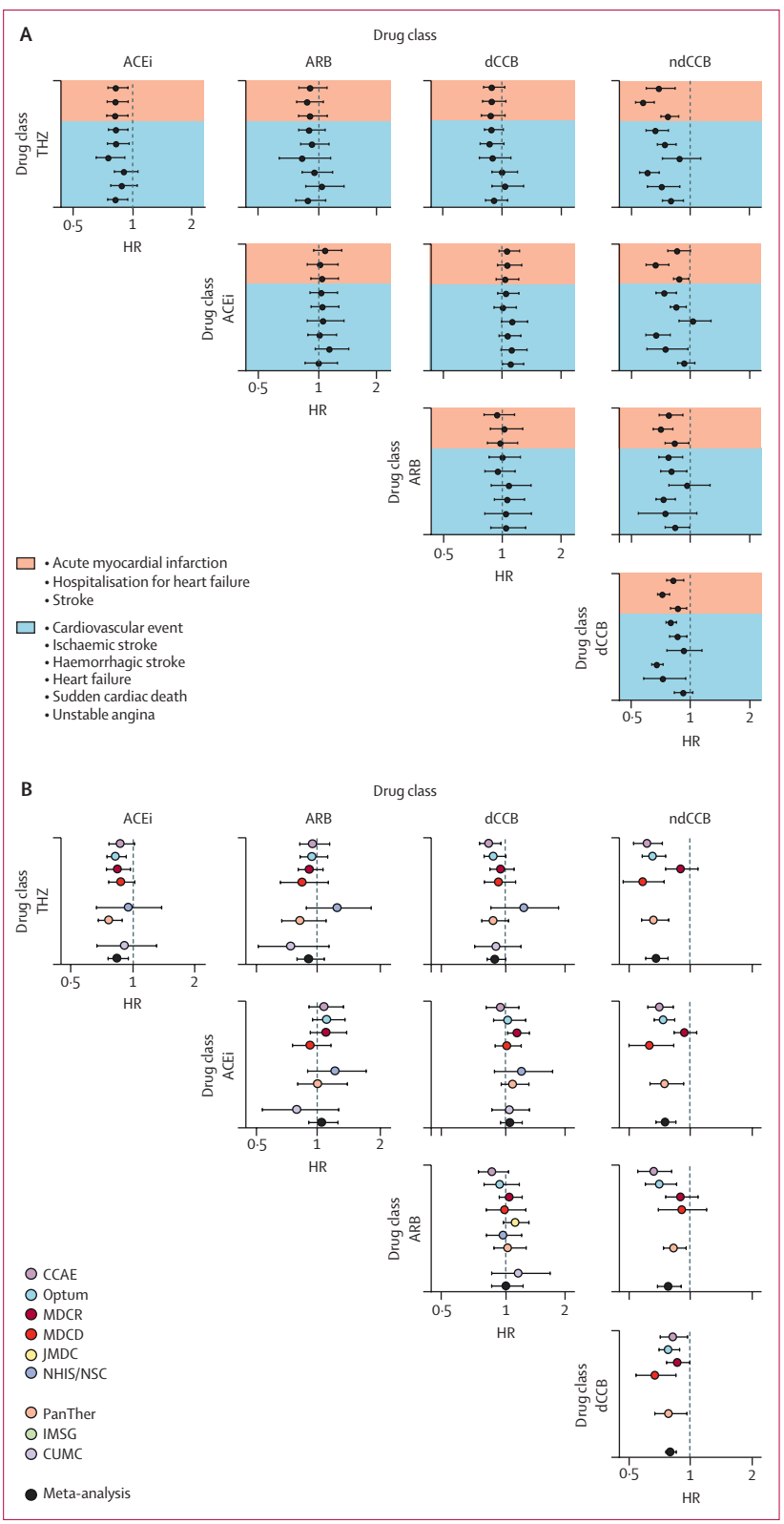

Figure 1: **Comparative effectiveness of THZs, ACEis, ARBs, dCCBs, and ndCCBs**
Points report hazard ratio (HR) estimates and lines mark their 95% CIs. HRs of less than 1 favour target (row) over comparator (column). (A) Meta-analytic risk estimates across all nine effectiveness outcomes with primary outcomes in orange and secondary outcomes in blue. (B) Cardiovascular event risk estimates by data source and meta-analysis. Colours identify databases; the top block are administrative claims databases, the middle block are electronic health records and black highlights a meta-analysis across all other sources. Not all databases contain sufficient new users for study inclusion. Cardiovascular event is a composite outcome of acute myocardial infarction, hospitalisation for heart failure, stroke, and sudden cardiac death. ACEi=angiotensin converting-enzyme inhibitors. ARB=angiotensin receptor blockers. dCCB=dihydropyridine calcium channel blockers. ndCCB=non-dihydropyridine calcium channel blockers. THZ=thiazide or thiazide-like diuretics. HR=hazard ratio. CCAE=IBM MarketScan Commercial Claims and Encounters. Optum=Optum ClinFormatics. MDCR=IBM MarketScan Medicare Supplemental Beneficiaries. MDCD=IBM MarketScan Multi-state Medicaid. JMDC=Japan Medical Data Center. NHIS/NSC=South Korea National Health Insurance Service/National Sample Cohort. PanTher=Optum Pan-Therapeutic. IMSG=IMS/IQVIA Disease Analyzer Germany. CUMC=Columbia University Medical Center.

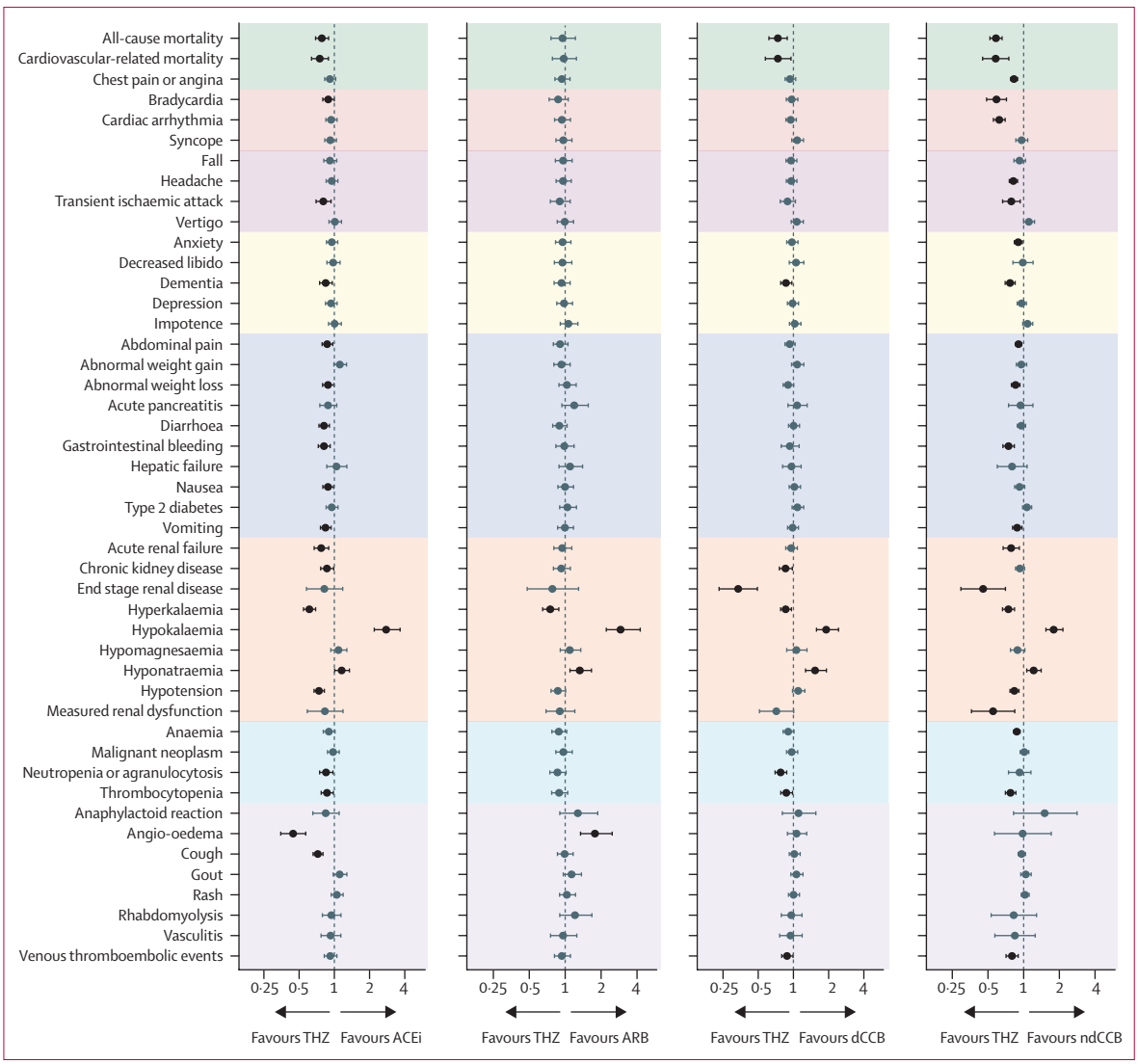

Figure 2: **Meta-analytic safety profiles comparing THZ to ACEi, ARB, dCCB, and ndCCB new users across 46 outcomes listed on product labels**
Points and lines identify HR estimates with their 95% CIs, respectively. Outcomes in grey signify that the CI covers HR of 1 (null hypothesis of no differential risk). THZ=thiazide or thiazide-like diuretics. ACEi=angiotensin converting-enzyme inhibitors. ARB=angiotensin receptor blockers. dCCB=dihydropyridine calcium channel blockers. ndCCB=non-dihydropyridine calcium channel blockers. HR=hazard ratio.

The effect of adjusting for baseline blood pressure across all nine effectiveness outcomes for all class pairs in the Optum Pan-Therapeutic database is presented in figure 3. Of 90 HR estimates, only three cases change their statistically significant interpretation when incorporating blood pressure in the propensity score model. The risk of acute myocardial infarction in thiazide or thiazide-like diuretic versus angiotensin-converting enzyme inhibitors new users moved from HR 0·81 (95% CI 0·68–0·98) to 0·85 (0·70–1·03) and the 95% CIs measuring the risk of acute myocardial infarction and stroke in dihydropyridine calcium channel blockers and non-dihydropyridine calcium channel blockers no longer covered an HR of 1. Similar consistency was seen between estimates for the safety profile of thiazide or thiazide-like diuretics versus angiotensin-converting enzyme inhibitors (appendix p 82).

Discussion

LEGEND-HTN is, to our knowledge, the largest and most comprehensive study to provide evidence about the comparative effectiveness and safety of first-line antihypertensives, representing 4·9 million patients initiating monotherapy across nine databases from four countries, examining all pairwise comparisons between the five first-line drug classes against a panel of 55 health outcomes. This equates to 22 000 traditional observational studies, many of which researchers could have hand-picked, tweaked, and published individually. Most comparisons of effectiveness revealed no differences

between classes. We found, however, that patients initiating treatment with a thiazide or thiazide-like diuretic had a significantly lower risk of seven effectiveness outcomes, including acute myocardial infarction, hospitalisation for heart failure, and stroke, as compared with angiotensin-converting enzyme inhibitors new users, while patients remain on-treatment with their initial drug class choice. Additionally, the thiazide or thiazide-like diuretic safety profile was markedly better compared with angiotensin-converting enzyme inhibitors. Patients who initiated treatment with a non-dihydropyridine calcium channel blocker had a significantly higher risk of poor effectiveness outcomes compared with all other class choices, but less adequate cohort balance and equipoise in these comparisons might limit their generalisability. Finally, there were no significant effectiveness differences between the remaining classes.

Across the patients who initiated monotherapy, nearly 50% were prescribed angiotensin-converting enzyme inhibitors and fewer than 18% thiazide or thiazide-like diuretics. Although our results suggest angiotensin-converting enzyme inhibitors are only modestly less effective than thiazide or thiazide-like diuretics, the effect of monotherapy with thiazide or thiazide-like diuretics across the whole population could be substantial; if the 2·4 million angiotensin-converting enzyme inhibitors new users had instead started on a thiazide or thiazide-like diuretic, more than 3100 major cardiovascular events could potentially have been avoided. This number equates to 1·3 cardiovascular events avoided for every 1000 patients who initiate with a thiazide or thiazide-like diuretic instead of an angiotensin-converting enzyme inhibitor, yielding a substantial public health impact, particularly given the more favourable safety profile of thiazide or thiazide-like diuretics.

Real-world observational studies can fill evidence gaps from what can be learned from RCTs. Whereas RCTs remain a key tool for high-quality clinical efficacy estimates in patient-limited, controlled settings, LEGEND-HTN delivers estimates of real-world effectiveness.[20] For example, the 2017 ACC/AHA Blood Pressure Treatment Guideline systematic review did a meta-analysis of three RCTs[3,21,22] to estimate the relative risk (RR) of myocardial infarction between 18421 thiazide or thiazide-like diuretic and 12225 angiotensin-converting enzyme inhibitors users in total, yielding an RR of 1·2 (95% CI 0·78–2·0). This estimate is concordant with, but lacks the statistical power of the LEGEND-HTN estimate involving more than 2·2 million patients with greater real-world heterogeneity. We note that the LEGEND-HTN estimate of myocardial infarction risk is not concordant with any of the three individual RCTs, but their marked differences with each other leaves the question of risk unanswered. For important efficacy outcomes, head-to-head RCTs between specific drug classes (eg, thiazide or thiazide-like diuretic *vs* angiotensin receptor blocker for risk of heart failure, and thiazide or thiazide-like diuretic *vs* angiotensin receptor blocker, and angiotensin-converting enzyme inhibitors *vs* angiotensin receptor blocker for risk of major cardiovascular events and renal events) do not exist.[4] For convenience, extant RCTs usually recruit previously treated hypertensive patients; LEGEND-HTN, on the other hand, focuses on treatment initiation and so directly assesses initiation guidelines. Finally, although RCTs and the systematic review furnish a comprehensive summary of cardiovascular outcomes, little evidence exists about the comparative safety of these classes. LEGEND-HTN provides these measures for all class comparisons.

Through an international network, LEGEND-HTN seeks to take advantage of disparate health databases drawn from different sources and across a range of countries and practice settings. These large-scale and unfiltered populations better represent real-world practice than the restricted study populations in prescribed treatment and follow-up settings from RCTs. The strong agreement among the separate database estimates despite heterogeneity in patient populations, practice settings, and data-capture processes further supports the plausibility of true causal effect differences. Even with

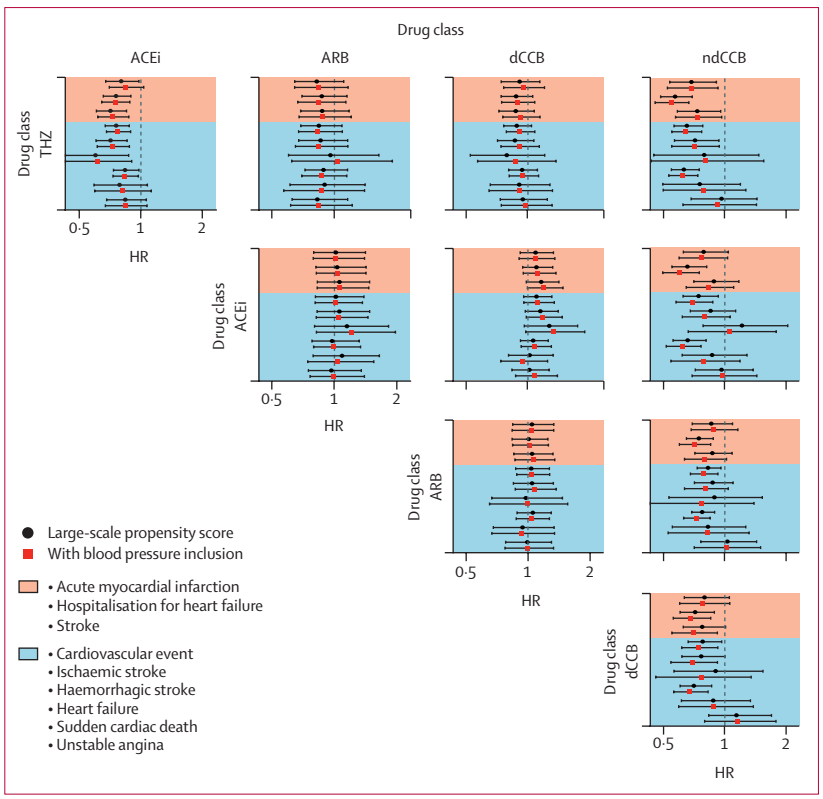

***Figure 3*: Effectiveness estimates comparing THZ to ACEi, ARB, dCCB, and ndCCB new users using propensity scores with and without baseline blood pressure adjustment in the Optum Pan-Therapeutic database**
Points and lines identify HR estimates with their 95% CIs, respectively. Black circles demarcate estimates based on large-scale propensity scores built without blood pressure measurements and red squares identify estimates additionally including baseline measurements from the electronic health record. ACEi=angiotensin converting-enzyme inhibitors. ARB=angiotensin receptor blockers. dCCB=dihydropyridine calcium channel blockers. ndCCB=non-dihydropyridine calcium channel blockers. THZ=thiazide or thiazide-like diuretics. HR=hazard ratio.

this greater generalisability, we cannot exclude the possibility of subpopulations not sufficiently captured in our research network that feature a considerably different effectiveness profile.

A limitation of LEGEND-HTN is the absence of blood pressure measurements within some databases. Baseline blood pressure might drive class choice, resulting in unmeasured confounding by indication between cohorts. For example, physicians might preferentially prescribe a thiazide or thiazide-like diuretic rather than an angiotensin-converting enzyme inhibitor for patients with low baseline blood pressure. If uncorrected, this can bias risk estimates to favour thiazide or thiazide-like diuretics given the strong correlation between high blood pressure and cardiovascular events. In Optum Pan-Therapeutic, however, we observed that thiazide or thiazide-like diuretic new users have the highest median blood pressure across drug classes. Unfortunately, there is no guarantee that this relationship holds in other data sources. To protect against such confounding, LEGEND-HTN used large-scale propensity score models involving tens of thousands of baseline patient characteristics, many of which should also associate with blood pressure to facilitate its indirect adjustment in spite of remaining unobserved. A post-hoc sensitivity analysis revealed that including blood pressure in the propensity score model achieved near-perfect balance on baseline blood pressure across comparisons in Optum Pan-Therapeutic, but did not lead to clinically meaningfully different effect sizes estimates than when not including blood pressure.

The standardisation in LEGEND-HTN enabled us to consider multiple study design choices. One choice was the time-at-risk definition. On-treatment time results in shorter follow-up than intention to treat. As expected, we saw blunted estimates of differential effectiveness and risks between drug class new users under an intention-to-treat design (appendix pp 57–68). We caution against overinterpretation of estimate differences between time-at-risk choices, because treatment escalation is more likely to confound intention-to-treat estimates.

On-treatment follow-up also helps to assess differential adherence to initial treatment. Except in the Columbia University Medical Center database, median on-treatment time was modestly shorter (0–38 days) for thiazide or thiazide-like diuretics versus angiotensin-converting enzyme inhibitors new users. Such differences, if meaningful, are also less likely to confound on-treatment estimates where time-at-risk ends with treatment discontinuation. Further, claims databases reported drug fulfilment whereas electronic health records reported prescriptions. Because fulfilment more directly reflects actual drug taking, one might expect differential adherence to generate notable effect estimate differences across data sources; we did not observe such differences in comparing thiazide or thiazide-like diuretics versus angiotensin-converting enzyme inhibitors new users.

Finally, cardiovascular observational research has a poor track record when it comes to reliability and reproducibility.[23] One probable cause is residual confounding due to the observational nature of the studies. In contrast to most observational research, LEGEND-HTN minimises the risk of residual bias by using reproducible methods to address observed confounding, by reporting study diagnostics such as empirical equipoise and covariate balance, and by unprecedentedly applying a large set of control outcomes to measure and then account for remaining systematic error. Marked covariate balance and empirical equipoise between new-user cohorts across data sources show here successful adjustment for observed confounding and comparable, generalisable populations for HR estimation. Control experiments further reduce systematic error and return calibrated CIs and p values with reliable statistical interpretation. Other causes of concern are publication bias and p-hacking that LEGEND-HTN addresses by consistently applying our study design to many comparisons and reporting all results through its interactive website. This further enables multiple testing correction as appropriate. Finally, LEGEND-HTN delivers true open science, with all study artifacts including study protocol, analytical code, and full results made publicly available. As a consequence, LEGEND-HTN evidence should show high reliability.[24]

Contributors

MAS, MJS, SCY, NP, DM, GH, and PBR contributed to study design, implementation and execution, data analysis and interpretation, and paper writing. HMK, RC, CGR, and JD contributed to data interpretation and paper writing.

Declaration of interests

PBR and MJS are employees of Janssen Research & Development, a subsidiary of Johnson & Johnson. CGR is an employee of IQVIA, whose customers are the entire pharmaceutical industry, among which are the manufacturers of the studied drugs. GH and MAS have received grant funding from Janssen to support methods research not directly related to this study. Neither Janssen nor IQVIA had input in the design, execution, interpretation of results or decision to publish. DM reports personal fees from Simon Greenstone Panatier, Williams Hart, Lieff Cabraser, and the Lanier Law firms. HMK reports personal fees from UnitedHealth, IBM Watson Health, Element Science, Aetna, Facebook, Arnold & Porter, Siegfied & Jensen, and the Ben C Martin Law Firms; grants from the Centers for Medicare & Medicaid Services, Medtronic, Johnson & Johnson, and the US Food and Drug Administration; and serving as founder of the personal health information platform Hugo and cofounder of Refactor Health, a health-care artificial intelligence-augmented data management company, outside the submitted work. All other authors declare no competing interests.

Acknowledgments

This work is partially supported through the US National Science Foundation (grant IIS 1251151), US National Institutes of Health grants (U19 AI135995 and R01 LM006910), South Korean Ministry of Health & Welfare (grant H116C0992), and the Australian National Health and Medical Research Council (grant GNT1157506).

References

1 Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2018; **138:** e426–83.

2 Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH guidelines for the management of arterial hypertension. *Eur Heart J* 2018; **39:** 3021–104.

3 The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). *JAMA* 2002; **288:** 2981–97.

4 Reboussin DM, Allen NB, Griswold ME, et al. Systematic review for the 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2018; **138:** e595–616.

5 Musini VM, Gueyffier F, Puil L, Salzwedel DM, Wright JM. Pharmacotherapy for hypertension in adults aged 18 to 59 years. *Cochrane Database Syst Rev* 2017; **8:** CD008276.

6 Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015; **216:** 574–78.

7 Ryan PB, Schuemie MJ, Gruber S, Zorych I, Madigan D. Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. *Drug Saf* 2013; **36** (suppl 1)**:** S59–72.

8 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016; **183:** 758–64.

9 ACCORD Study Group, Cushman WC, Evans GW, et al. Effects of intensive blood-pressure control in type 2 diabetes mellitus. *N Engl J Med* 2010; **362:** 1575–85.

10 SPRINT Research Group, Wright JT Jr, Williamson JD, et al. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med* 2015; **373:** 2103–16.

11 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70:** 41.

12 Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int J Epidemiol* 2018; **47:** 2005–14.

13 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7:** 177–88.

14 Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med* 2013; **33:** 209–18.

15 Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Robust empirical calibration of p-values using observational data. *Stat Med* 2016; **35:** 3883–88.

16 Voss EA, Boyce RD, Ryan PB, van der Lei J, Rijnbeek PR, Schuemie MJ. Accuracy of an automated knowledge base for identifying drug adverse reactions. *J Biomed Inform* 2017; **66:** 72–81.

17 Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc Natl Acad Sci USA* 2018; **115:** 2571–77.

18 Walker A, Patrick, Lauer, et al. A tool for assessing the feasibility of comparative effectiveness research. *Comparative Effectiveness Research* 2013; **3:** 112–20.

19 Suchard MA, Simpson SE, Zorych I, Ryan P, Madigan D. Massive parallelization of serial inference algorithms for a complex generalized linear Model. *ACM Trans Model Comput Simul* 2013; **23:** 1–17.

20 Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998; **316:** 285.

21 Zanchetti A, Crepaldi G, Bond MG, et al. Different effects of antihypertensive regimens based on fosinopril or hydrochlorothiazide with or without lipid lowering by pravastatin on progression of asymptomatic carotid atherosclerosis: principal results of PHYLLIS—a randomized double-blind trial. *Stroke* 2004; **35:** 2807–12.

22 Wing LMH, Reid CM, Ryan P, et al. A comparison of outcomes with angiotensin-converting-enzyme inhibitors and diuretics for hypertension in the elderly. *N Engl J Med* 2003; **348:** 583–92.

23 Rush CJ, Campbell RT, Jhund PS, Petrie MC, McMurray JJV. Association is not causation: treatment effects cannot be estimated from observational data in heart failure. *Eur Heart J* 2018; **39:** 3417–38.

24 Schuemie MJ, Ryan PB, Hripcsak G, Madigan D, Suchard MA. Improving reproducibility by using high-throughput observational studies with empirical calibration. *Philos Trans A Math Phys Eng Sci* 2018; **376:** 20170356.