RESEARCH PROTOCOL

Quantitative bias analysis for outcome phenotype error correction in comparative effect estimation: an empirical and simulation study

Table of Contents

1 Investigators

Investigator	Contact
James Weaver [1,2,3]	james.weaver@ndorms.ox.ac.uk
Patrick B. Ryan [2,3,4]	ryan@ohdsi.org
Victoria Strauss [1,2]	victoria.strauss@csm.ox.ac.uk
Marc A. Suchard [3,5]	msuchard@ucla.edu
Joel Swerdel [2,3]	jswerdel@its.jnj.com
Daniel Prieto-Alhambra [1,3,6]	daniel.prietoalhambra@ndorms.ox.ac.uk

[1]Centre for Statistics in Medicine, Nuffield Department of Orthopedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK; [2]Observational Health Data Analytics, Janssen Research and Development, Titusville, NJ, USA; [3]Observational Health Data Sciences and Informatics, New York, NY, USA; [4]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA; [5]Department of Biostatistics, Fielding School of Public Health, and Department of Biomathematics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA; [6]Medical Informatics, Erasmus Medical Centre, Rotterdam, Netherlands

1.1 Disclosures

This study is undertaken within Observational Health Data Sciences and Informatics (OHDSI), an open science collaboration. **JW**, **PR**, and **JS** are employees of shareholders of Janssen R&D (a Johnson & Johnson company). **MAS** receives grant support from the US National Institutes of Health, US Food & Drug Administration and US Department of Veterans Affairs and contracts from Janssen R&D. **VS** has no conflicts of interest to declare. **DPA**'s research group has received grant support from Amgen, Chesi-Taylor, Novartis, and UCB Biopharma. His department has received advisory, consultancy fees from Amgen, Astellas, AstraZeneca, Johnson & Johnson, and UCB Biopharma and fees for speaker services from Amgen and UCB Biopharma. Janssen, on behalf of IMI-funded EHDEN and EMIF consortiums, and Synapse Management Partners have supported training programs organized by DPA's department and open for external participants organized by his department outside submitted work.

2 Rationale and background

Phenotype error is acknowledged but rarely corrected for in causal effect estimation studies using observational data. Quantitative bias analysis (QBA) is a method for phenotype error correction, but the extent to which it minimizes bias in effect estimates is unclear.

3 Study objectives

- Empirically evaluate QBA for outcome phenotype error correction in several pharmacoepidemiologic comparative effect estimation scenarios

- Simulate an analytic space defined by outcome incidence proportions (IP), observed effect estimates, and phenotype measurement errors to determine which QBA input combinations produce valid results.

4 Research methods

4.1 Quantitative bias analysis

QBA is a method for correcting outcome phenotype error that can bias comparative effect estimates[1,2]. Phenotype definition performance characteristics (sensitivity and specificity, positive and negative predictive value) are required inputs for applying QBA. We use QBA for outcome phenotype error correction per the equations in **Table 1a** and **Table 1b**.

4.1.1 Table 1a: Observed exposure by outcome contingency table

Outcome	T	C
O[+]	a	b
O[-]	c	d
Total	a + c	a + d

T: target exposure, C: comparator exposure

4.1.2 Table 1b Expected exposure by outcome contingency table corrected for outcome phenotype error

Outcome	T	C
O[+]	A = a - (1 - SP1) * (a + c)) / (SN1 - (1 - SP1)	B = b - (1 - SP0) * (b + d)) / (SN0 - (1 - SP0)
O[-]	C = (a + c) - A	D = (b + d) - B
Total	A + C	B + D

T: target exposure, C: comparator exposure, SP1: specificity in target cohort, SN1: sensitivity in target cohort, SP0: specificity in comparator cohort, SN0: sensitivity in comparator cohort

4.2 Probabilistic reference standard validation

PheValuator is a method to calculate the performance characteristics of phenotype algorithms, namely, sensitivity, specificity, and positive and negative predictive value[3,4]. It develops a diagnostic predictive model to determine a probabilistic reference standard of patients against which phenotype algorithm performance can be assessed. **Table 2** reports the PheValuator confusion matrix and error metric calculations.

4.2.1 Table 2: PheValuator confusion matrix

Diagnostic model output	Phenotype algorithm case	Phenotype algorithm non-case
Predicted probability from diagnostic predictive model, P(Y)	TP = Σ[P(Y | Case)]	FP = Σ[1 - P(Y | Case)]
Predicted probability from diagnostic predictive model, P(Y)	FN = Σ[P(Y | Non-case)]	TN = Σ[1 - P(Y | Non-case)]

The following cohorts are required input for a PheValuator validation study:

- Ischemic stroke events during inpatient or emergency room visits
- Extremely specific cohort (xSpec)
- Extremely sensitive cohort (xSens)
- Prevalence cohort
- Database population evaluation cohort
- Exposure population evaluation cohorts
  - Note, this is a cohort shell with placeholders where drug exposure and condition occurrence concept sets. These replacements are made to construct the following exposure population evaluation cohorts:
    - ACE exposed without ischemic stroke OR ACE exposed with subsequent ischemic stroke
    - ARB exposed without ischemic stroke OR ARB exposed with subsequent ischemic stroke
    - THZ exposed without ischemic stroke OR THZ exposed with subsequent ischemic stroke

Detailed cohort definitions for probabilistic reference standard validation are available in Appendix Section A.

4.3 Empirical example

4.3.1 Design

Active comparator, new user comparative cohort study to estimate the risk of ischemic stroke among patients with hypertension initiating:

- Angiotensin-converting enzyme inhibitors (ACE) vs angiotensin receptor blockers (ARB)
- Angiotensin-converting enzyme inhibitors (ACE) vs Thiazide/thiazide-like diuretics (THZ)

4.3.2 Exposure cohort definitions

Detailed exposure cohort definitions for 3 class-level hypertension treatments are in Appendix Section B.

4.3.2.1 ACEI new users with prior hypertension

- First use of ACEI on or after January 1, 2010 with ≥365 days of prior continuous database observation
  - ≥1 condition occurrence of hypertension between 365 and 0 days relative to first use
  - exactly 1 exposure to hypertension medications between 0 and 7 days relative to first use
  - no prior exposure to hypertension medications

4.3.2.2 ARB new users with prior hypertension

- First use of ARB on or after January 1, 2010 with ≥365 days of prior continuous database observation
  - ≥1 condition occurrence of hypertension between 365 and 0 days relative to first use
  - exactly 1 exposure to hypertension medications between 0 and 7 days relative to first use
  - no prior exposure to hypertension medications

4.3.2.3 THZ new users with prior hypertension

- First use of THZ on or after January 1, 2010 with ≥365 days of prior continuous database observation
  - ≥1 condition occurrence of hypertension between 365 and 0 days relative to first use
  - exactly 1 exposure to hypertension medications between 0 and 7 days relative to first use
  - no prior exposure to hypertension medications

4.3.3 Outcome definition

- Inpatient or emergency room visits on or after January 1, 2010
  - ≥1 condition occurrence of ischemic stroke
  - exactly 0 condition occurrences of ischemic stroke between -365 and -1 days relative to inpatient or emergency room visit with ischemic stroke

The detailed outcome definition for inpatient ischemic stroke is in Appendix Section C.

4.3.4 Data sources

The study will be executed against 4 US adminstrative healthcare claims and 1 US electronic health record databases.

- Optum® de-identified Clinformatics® Datamart - Date of Death (optum_extended_dod)
- Optum® Electronic Health Record (optum_ehr)
- IBM MarketScan® Commercial Database (truven_ccae)
- IBM MarketScan® Multi-State Medicaid (truven_mdcd)

- IBM MarketScan® Medicare Supplemental Beneficiaries (truven_mdcr)

The database descriptions are in Appendix Section D.

4.3.5 Outcome definition

See Appendix Section B for detailed exposure definitions.

4.3.6 Time-at-risk

- 1 day to 365 days relative to exposure start
- 1 day to 730 days relative to exposure start

4.3.7 Analyses

- Calculate database-level (i.e., non-differential) and exposure-level (i.e., differential) ischemic stroke phenotype definition sensitivity and specificity using probabilistic reference standard validation studies in each data source
- Estimate comparative treatment effect using logistic regression (odds ratio [OR] with 95% confidence intervals [CI]) for ACE vs ARB and ACE vs THZ under the following analysis specifications:
  - Unadjusted
  - Non-differential QBA adjustment
  - Differential QBA adjustment
  - 1:1 propensity score (PS) matched
  - 1:1 PS matched with non-differential QBA
  - 1:1 PS matched with differential QBA
- Execute 5 databases x 2 comparisons x 2 TARs x 6 analyses] = 120 analyses

4.3.8 Evaluation metrics

QBA performance evaluated by bias difference, relative bias, squared error, and precision difference between analyses that did vs did not include QBA.

- **Bias difference:** $\log(OR) - \log(OR_{QBA})$
- **Relative bias:** $(OR - OR_{QBA}) / OR * 100$
- **Relative precision:** $(1 / (SE(\log(OR))^2) - (1 / (SE(\log(OR_{QBA}))^2) / (1 / (SE(\log(OR))^2) * 100$
- **Squared error:** $(\log(OR) - \log(OR_{QBA})^2$

4.4 Grid space simulation

4.4.1 Inputs

Create grid space of all combinations of 4 input parameters:

- 5 outcome incidence proportions (IP) [10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5}]
- 6 uncorrected odds ratios (OR) [1, 1.25, 1.50, 2, 4, 10]

- 20 non-differential sensitivity values [0.05 to 1.00 by 0.05]
- 20 non-differential specificity values [1 - prevalence to 1.00 by 5%^{ile}]

The complete grid space consists of 12,000 2x2 contingency tables, each with 1,000,000 target and 1,000,000 comparator exposures and associated inputs.

4.4.2 Analysis

For each IP-OR combination, compute a distribution of QBA-corrected ORs with 95% CIs across combinations of sensitivity and specificity values and plot contours across the complete IP by OR grid space.

4.4.3 Evaluation metrics

The grid space simulation analysis will be evaluated by bias difference and relative bias between the unadjusted OR and the 25%^{ile}, 50%^{ile}, 75%^{ile}, and maximum of the QBA-corrected distribution of estimates. Report the overall, IP-stratified, OR-stratified, and IP-OR-stratified proportion of the total grid space that produces valid (i.e., non-zero) QBA-corrected counts.

- **Bias difference:** $\log(OR) - \log(OR_{QBA})$
- **Relative bias:** $(OR - OR_{QBA}) / OR * 100$

5 Strengths and limitations

5.1 Strengths

- Empirical evaluation strictly systematic using common data model and standardized analytic tools

5.2 Limitations

- Empirical example uses simple and multidimensional QBA only, no probabilistic QBA or multiple bias modeling
- Only uses sensitivity and specificity approach, no use of PPV and NPV
- Logistic regression outcome model assumes constant risk, discards survival information
- Assumes probabilistic reference validation metrics are accurate
- Validation study within exposure-indication populations will have incomplete overlap with restricted study populations

6 Protection of human subjects

This work does not involve human patient research. It uses de-identified patient-level data collected during routine healthcare provision. Confidentiality of patient records will be

maintained. Study reports will contain aggregate data only and will not identify individual patients of care providers.

References

1 Lash TL, Fox MP, Fink AK. *Applying quantitative bias analysis to epidemiologic data*. Springer New York 2009.

2 Lash TL, Fox MP, MacLehose RF, *et al.* Good practices for quantitative bias analysis. *Int J Epidemiol* 2014;**43**:1969–85.

3 Swerdel JN, Hripcsak G, Ryan PB. PheValuator: Development and evaluation of a phenotype algorithm evaluator. *J Biomed Inform* 2019;**97**:103258.

4 Swerdel JN, Schuemie M, Murray G, *et al.* PheValuator 2.0: Methodological improvements for the PheValuator approach to semi-automated phenotype algorithm evaluation. *J Biomed Inform* 2022;**135**:104177.

Appendix

7 Probabilistic reference standard validation cohort definitions

7.1 xSpec validation cohort

7.1.1 Cohort Entry Events

People may enter the cohort when observing any of the following:

1. visit occurrences of any visit, starting on or after January 1, 2010.

Restrict entry events to having at least 2 condition occurrences of '[QBA eval] Cerebral infarction NC', starting in the 1 days prior to cohort entry start date.

7.1.2 Cohort Exit

The cohort end date will be offset from index event's start date plus 365 days.

7.1.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

7.2 xSens validation cohort

7.2.1 Cohort Entry Events

People may enter the cohort when observing any of the following:

1. visit occurrences of any visit, starting on or after January 1, 2010.

Restrict entry events to having at least 1 condition occurrence of '[QBA eval] Cerebral infarction NC', starting in the 91 days prior to cohort entry start date.

7.2.2 Cohort Exit

The cohort end date will be offset from index event's start date plus 365 days.

7.2.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

7.3 Prevalence validation cohort

7.3.1 Cohort Entry Events

People enter the cohort when observing any of the following:

1. condition occurrences of '[QBA eval] Cerebral infarction NC', starting on or after January 1, 2010.

Limit cohort entry events to the earliest event per person.

7.3.2 Cohort Exit

The person also exists the cohort at the end of continuous observation.

7.3.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

7.4 Database population evaluation cohort

7.4.1 Cohort Entry Events

People enter the cohort when observing any of the following:

1. visit occurrences of 'Inpatient or Inpatient/ER visit', starting on or after January 1, 2010; having no condition occurrences of '[QBA eval] Cerebral infarction NC'.

2. visit occurrences of 'Inpatient or Inpatient/ER visit', starting on or after January 1, 2010; having at least 1 condition occurrence of '[QBA eval] Cerebral infarction NC', starting between 0 days before and all days after 'Inpatient or Inpatient/ER visit' start date and starting anytime on or before 'Inpatient or Inpatient/ER visit' end date.

7.4.2 Cohort Exit

The cohort end date will be offset from index event's start date plus 365 days.

7.4.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

7.5 ACEI new users evaluation cohort

7.5.1 Cohort Entry Events

People enter the cohort when observing any of the following:

1. visit occurrences of 'Inpatient or Inpatient/ER visit', starting on or after January 1, 2010; with all of the following criteria:

2. having no condition occurrences of '[QBA eval] Cerebral infarction NC'.

3. having at least 1 drug exposure of '[QBA eval] ACEIs', starting anytime on or before 'Inpatient or Inpatient/ER visit' start date; having at least 1 condition occurrence of '[QBA eval] hypertension', starting anytime on or before '[QBA eval] ACEIs' start date.

4. visit occurrences of 'Inpatient or Inpatient/ER visit', starting on or after January 1, 2010; with all of the following criteria:

5. having at least 1 condition occurrence of '[QBA eval] Cerebral infarction NC', starting between 0 days before and all days after 'Inpatient or Inpatient/ER visit' start date and starting anytime on or before 'Inpatient or Inpatient/ER visit' end date.

6. having at least 1 drug exposure of '[QBA eval] ACEIs', starting anytime on or before 'Inpatient or Inpatient/ER visit' start date; having at least 1 condition occurrence of '[QBA eval] hypertension'.

7.5.2 Cohort Exit

The cohort end date will be offset from index event's start date plus 365 days.

7.5.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

7.6 ARB new users evaluation cohort

7.6.1 Cohort Entry Events

People enter the cohort when observing any of the following:

1. visit occurrences of 'Inpatient or Inpatient/ER visit', starting on or after January 1, 2010; with all of the following criteria:

2. having no condition occurrences of '[QBA eval] Cerebral infarction NC'.

3. having at least 1 drug exposure of '[QBA eval] ARBs', starting anytime on or before 'Inpatient or Inpatient/ER visit' start date; having at least 1 condition occurrence of '[QBA eval] hypertension', starting anytime on or before '[QBA eval] ARBs' start date.

4. visit occurrences of 'Inpatient or Inpatient/ER visit', starting on or after January 1, 2010; with all of the following criteria:

5. having at least 1 condition occurrence of '[QBA eval] Cerebral infarction NC', starting between 0 days before and all days after 'Inpatient or Inpatient/ER visit' start date and starting anytime on or before 'Inpatient or Inpatient/ER visit' end date.

6. having at least 1 drug exposure of '[QBA eval] ARBs', starting anytime on or before 'Inpatient or Inpatient/ER visit' start date; having at least 1 condition occurrence of '[QBA eval] hypertension'.

7.6.2 Cohort Exit

The cohort end date will be offset from index event's start date plus 365 days.

7.6.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

7.7 THS new users evaluation cohort

7.7.1 Cohort Entry Events

People enter the cohort when observing any of the following:

1. visit occurrences of 'Inpatient or Inpatient/ER visit', starting on or after January 1, 2010; with all of the following criteria:

2. having no condition occurrences of '[QBA eval] Cerebral infarction NC'.

3. having at least 1 drug exposure of '[QBA eval] THZs', starting anytime on or before 'Inpatient or Inpatient/ER visit' start date; having at least 1 condition occurrence of '[QBA eval] hypertension', starting anytime on or before '[QBA eval] THZs' start date.

4. visit occurrences of 'Inpatient or Inpatient/ER visit', starting on or after January 1, 2010; with all of the following criteria:

5. having at least 1 condition occurrence of '[QBA eval] Cerebral infarction NC', starting between 0 days before and all days after 'Inpatient or Inpatient/ER visit' start date and starting anytime on or before 'Inpatient or Inpatient/ER visit' end date.

6. having at least 1 drug exposure of '[QBA eval] THZs', starting anytime on or before 'Inpatient or Inpatient/ER visit' start date; having at least 1 condition occurrence of '[QBA eval] hypertension'.

7.7.2 Cohort Exit

The cohort end date will be offset from index event's start date plus 365 days.

7.7.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

---

8 Exposure cohort definitions

8.1 ACEI new users with prior hypertension

8.1.1 Cohort Entry Events

People with continuous observation of 365 days before event may enter the cohort when observing any of the following:

1. drug exposure of '[QBA eval] ACEIs' for the first time in the person's history, starting on or after January 1, 2010.

Limit cohort entry events to the earliest event per person.

Restrict entry events to with all of the following criteria:

1. having no drug exposures of '[QBA eval] hypertension drugs', starting anytime prior to cohort entry start date.
2. having at least 1 condition occurrence of '[QBA eval] hypertension', starting between 365 days before and 0 days after cohort entry start date.
3. having exactly 1 distinct standard concepts from drug era of '[QBA eval] hypertension drugs', starting between 0 days before and 7 days after cohort entry start date.

8.1.2 Cohort Exit

The cohort end date will be based on a continuous exposure to '[QBA eval] ACEIs': allowing 30 days between exposures, adding 0 days after exposure ends, and using days supply and exposure end date for exposure duration.

8.1.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

8.1.4 [QBA eval] hypertension

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
316866	Hypertensive disorder	38341003	SNOMED	NO	YES	NO

8.1.5 [QBA eval] hypertension drugs

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1319998	acebutolol	149	RxNorm	NO	YES	NO
1317967	aliskiren	325646	RxNorm	NO	YES	NO
991382	amiloride	644	RxNorm	NO	YES	NO
1332418	amlodipine	17767	RxNorm	NO	YES	NO
1314002	atenolol	1202	RxNorm	NO	YES	NO
40235485	azilsartan	1091643	RxNorm	NO	YES	NO
1335471	benazepril	18867	RxNorm	NO	YES	NO
1322081	betaxolol	1520	RxNorm	NO	YES	NO
1338005	bisoprolol	19484	RxNorm	NO	YES	NO
932745	bumetanide	1808	RxNorm	NO	YES	NO
1351557	candesartan	214354	RxNorm	NO	YES	NO
1340128	captopril	1998	RxNorm	NO	YES	NO
1346823	carvedilol	20352	RxNorm	NO	YES	NO
1395058	chlorthalidone	2409	RxNorm	NO	YES	NO

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1398937	clonidine	2599	RxNorm	NO	YES	NO
1328165	diltiazem	3443	RxNorm	NO	YES	NO
1363053	doxazosin	49276	RxNorm	NO	YES	NO
1341927	enalapril	3827	RxNorm	NO	YES	NO
1309799	eplerenone	298869	RxNorm	NO	YES	NO
1346686	eprosartan	83515	RxNorm	NO	YES	NO
1353776	felodipine	4316	RxNorm	NO	YES	NO
1363749	fosinopril	50166	RxNorm	NO	YES	NO
956874	furosemide	4603	RxNorm	NO	YES	NO
1344965	guanfacine	40114	RxNorm	NO	YES	NO
1373928	hydralazine	5470	RxNorm	NO	YES	NO
974166	hydrochlorothiazide	5487	RxNorm	NO	YES	NO
978555	indapamide	5764	RxNorm	NO	YES	NO
1347384	irbesartan	83818	RxNorm	NO	YES	NO
1326012	isradipine	33910	RxNorm	NO	YES	NO
1386957	labetalol	6185	RxNorm	NO	YES	NO
1308216	lisinopril	29046	RxNorm	NO	YES	NO
1367500	losartan	52175	RxNorm	NO	YES	NO
1305447	methyldopa	6876	RxNorm	NO	YES	NO
907013	metolazone	6916	RxNorm	NO	YES	NO
1307046	metoprolol	6918	RxNorm	NO	YES	NO
1309068	minoxidil	6984	RxNorm	NO	YES	NO
1310756	moexipril	30131	RxNorm	NO	YES	NO
1313200	nadolol	7226	RxNorm	NO	YES	NO
1314577	nebivolol	31555	RxNorm	NO	YES	NO
1318137	nicardipine	7396	RxNorm	NO	YES	NO
1318853	nifedipine	7417	RxNorm	NO	YES	NO
1319880	nisoldipine	7435	RxNorm	NO	YES	NO
40226742	olmesartan	321064	RxNorm	NO	YES	NO
1327978	penbutolol	7973	RxNorm	NO	YES	NO
1373225	perindopril	54552	RxNorm	NO	YES	NO
1345858	pindolol	8332	RxNorm	NO	YES	NO
1350489	prazosin	8629	RxNorm	NO	YES	NO

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1353766	propranolol	8787	RxNorm	NO	YES	NO
1331235	quinapril	35208	RxNorm	NO	YES	NO
1334456	ramipril	35296	RxNorm	NO	YES	NO
970250	spironolactone	9997	RxNorm	NO	YES	NO
1317640	telmisartan	73494	RxNorm	NO	YES	NO
1341238	terazosin	37798	RxNorm	NO	YES	NO
942350	torsemide	38413	RxNorm	NO	YES	NO
1342439	trandolapril	38454	RxNorm	NO	YES	NO
904542	triamterene	10763	RxNorm	NO	YES	NO
1308842	valsartan	69749	RxNorm	NO	YES	NO
1307863	verapamil	11170	RxNorm	NO	YES	NO

8.1.6 [QBA eval] ACEIs

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1335471	benazepril	18867	RxNorm	NO	YES	NO
1340128	captopril	1998	RxNorm	NO	YES	NO
1341927	enalapril	3827	RxNorm	NO	YES	NO
1363749	fosinopril	50166	RxNorm	NO	YES	NO
1308216	lisinopril	29046	RxNorm	NO	YES	NO
1310756	moexipril	30131	RxNorm	NO	YES	NO
1373225	perindopril	54552	RxNorm	NO	YES	NO
1331235	quinapril	35208	RxNorm	NO	YES	NO
1334456	ramipril	35296	RxNorm	NO	YES	NO
1342439	trandolapril	38454	RxNorm	NO	YES	NO

8.2 ARB new users with prior hypertension

8.2.1 Cohort Entry Events

People with continuous observation of 365 days before event may enter the cohort when observing any of the following:

1. drug exposure of '[QBA eval] ARBs' for the first time in the person's history, starting on or after January 1, 2010.

Limit cohort entry events to the earliest event per person.

Restrict entry events to with all of the following criteria:

1. having no drug exposures of '[QBA eval] hypertension drugs', starting anytime prior to cohort entry start date.
2. having at least 1 condition occurrence of '[QBA eval] hypertension', starting between 365 days before and 0 days after cohort entry start date.
3. having exactly 1 distinct standard concepts from drug era of '[QBA eval] hypertension drugs', starting between 0 days before and 7 days after cohort entry start date.

8.2.2 Cohort Exit

The cohort end date will be based on a continuous exposure to '[QBA eval] ARBs': allowing 30 days between exposures, adding 0 days after exposure ends, and using days supply and exposure end date for exposure duration.

8.2.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

8.2.4 [QBA eval] hypertension

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
316866	Hypertensive disorder	38341003	SNOMED	NO	YES	NO

8.2.5 [QBA eval] hypertension drugs

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1319998	acebutolol	149	RxNorm	NO	YES	NO
1317967	aliskiren	325646	RxNorm	NO	YES	NO
991382	amiloride	644	RxNorm	NO	YES	NO
1332418	amlodipine	17767	RxNorm	NO	YES	NO
1314002	atenolol	1202	RxNorm	NO	YES	NO
40235485	azilsartan	1091643	RxNorm	NO	YES	NO
1335471	benazepril	18867	RxNorm	NO	YES	NO
1322081	betaxolol	1520	RxNorm	NO	YES	NO
1338005	bisoprolol	19484	RxNorm	NO	YES	NO
932745	bumetanide	1808	RxNorm	NO	YES	NO
1351557	candesartan	214354	RxNorm	NO	YES	NO
1340128	captopril	1998	RxNorm	NO	YES	NO
1346823	carvedilol	20352	RxNorm	NO	YES	NO
1395058	chlorthalidone	2409	RxNorm	NO	YES	NO

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1398937	clonidine	2599	RxNorm	NO	YES	NO
1328165	diltiazem	3443	RxNorm	NO	YES	NO
1363053	doxazosin	49276	RxNorm	NO	YES	NO
1341927	enalapril	3827	RxNorm	NO	YES	NO
1309799	eplerenone	298869	RxNorm	NO	YES	NO
1346686	eprosartan	83515	RxNorm	NO	YES	NO
1353776	felodipine	4316	RxNorm	NO	YES	NO
1363749	fosinopril	50166	RxNorm	NO	YES	NO
956874	furosemide	4603	RxNorm	NO	YES	NO
1344965	guanfacine	40114	RxNorm	NO	YES	NO
1373928	hydralazine	5470	RxNorm	NO	YES	NO
974166	hydrochlorothiazide	5487	RxNorm	NO	YES	NO
978555	indapamide	5764	RxNorm	NO	YES	NO
1347384	irbesartan	83818	RxNorm	NO	YES	NO
1326012	isradipine	33910	RxNorm	NO	YES	NO
1386957	labetalol	6185	RxNorm	NO	YES	NO
1308216	lisinopril	29046	RxNorm	NO	YES	NO
1367500	losartan	52175	RxNorm	NO	YES	NO
1305447	methyldopa	6876	RxNorm	NO	YES	NO
907013	metolazone	6916	RxNorm	NO	YES	NO
1307046	metoprolol	6918	RxNorm	NO	YES	NO
1309068	minoxidil	6984	RxNorm	NO	YES	NO
1310756	moexipril	30131	RxNorm	NO	YES	NO
1313200	nadolol	7226	RxNorm	NO	YES	NO
1314577	nebivolol	31555	RxNorm	NO	YES	NO
1318137	nicardipine	7396	RxNorm	NO	YES	NO
1318853	nifedipine	7417	RxNorm	NO	YES	NO
1319880	nisoldipine	7435	RxNorm	NO	YES	NO
40226742	olmesartan	321064	RxNorm	NO	YES	NO
1327978	penbutolol	7973	RxNorm	NO	YES	NO
1373225	perindopril	54552	RxNorm	NO	YES	NO
1345858	pindolol	8332	RxNorm	NO	YES	NO
1350489	prazosin	8629	RxNorm	NO	YES	NO

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1353766	propranolol	8787	RxNorm	NO	YES	NO
1331235	quinapril	35208	RxNorm	NO	YES	NO
1334456	ramipril	35296	RxNorm	NO	YES	NO
970250	spironolactone	9997	RxNorm	NO	YES	NO
1317640	telmisartan	73494	RxNorm	NO	YES	NO
1341238	terazosin	37798	RxNorm	NO	YES	NO
942350	torsemide	38413	RxNorm	NO	YES	NO
1342439	trandolapril	38454	RxNorm	NO	YES	NO
904542	triamterene	10763	RxNorm	NO	YES	NO
1308842	valsartan	69749	RxNorm	NO	YES	NO
1307863	verapamil	11170	RxNorm	NO	YES	NO

8.2.6 [QBA eval] ARBs

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
40235485	azilsartan	1091643	RxNorm	NO	YES	NO
1351557	candesartan	214354	RxNorm	NO	YES	NO
1346686	eprosartan	83515	RxNorm	NO	YES	NO
1347384	irbesartan	83818	RxNorm	NO	YES	NO
1367500	losartan	52175	RxNorm	NO	YES	NO
40226742	olmesartan	321064	RxNorm	NO	YES	NO
1317640	telmisartan	73494	RxNorm	NO	YES	NO
1308842	valsartan	69749	RxNorm	NO	YES	NO

8.3 THZ new users with prior hypertension

8.3.1 Cohort Entry Events

People with continuous observation of 365 days before event may enter the cohort when observing any of the following:

1. drug exposure of '[QBA eval] THZs' for the first time in the person's history, starting on or after January 1, 2010.

Limit cohort entry events to the earliest event per person.

Restrict entry events to with all of the following criteria:

1. having no drug exposures of '[QBA eval] hypertension drugs', starting anytime prior to cohort entry start date.

2. having at least 1 condition occurrence of '[QBA eval] hypertension', starting between 365 days before and 0 days after cohort entry start date.
3. having exactly 1 distinct standard concepts from drug era of '[QBA eval] hypertension drugs', starting between 0 days before and 7 days after cohort entry start date.

8.3.2 Cohort Exit

The cohort end date will be based on a continuous exposure to '[QBA eval] THZs': allowing 30 days between exposures, adding 0 days after exposure ends, and using days supply and exposure end date for exposure duration.

8.3.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

8.3.4 [QBA eval] hypertension

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
316866	Hypertensive disorder	38341003	SNOMED	NO	YES	NO

8.3.5 [QBA eval] hypertension drugs

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1319998	acebutolol	149	RxNorm	NO	YES	NO
1317967	aliskiren	325646	RxNorm	NO	YES	NO
991382	amiloride	644	RxNorm	NO	YES	NO
1332418	amlodipine	17767	RxNorm	NO	YES	NO
1314002	atenolol	1202	RxNorm	NO	YES	NO
40235485	azilsartan	1091643	RxNorm	NO	YES	NO
1335471	benazepril	18867	RxNorm	NO	YES	NO
1322081	betaxolol	1520	RxNorm	NO	YES	NO
1338005	bisoprolol	19484	RxNorm	NO	YES	NO
932745	bumetanide	1808	RxNorm	NO	YES	NO
1351557	candesartan	214354	RxNorm	NO	YES	NO
1340128	captopril	1998	RxNorm	NO	YES	NO
1346823	carvedilol	20352	RxNorm	NO	YES	NO
1395058	chlorthalidone	2409	RxNorm	NO	YES	NO
1398937	clonidine	2599	RxNorm	NO	YES	NO
1328165	diltiazem	3443	RxNorm	NO	YES	NO

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1363053	doxazosin	49276	RxNorm	NO	YES	NO
1341927	enalapril	3827	RxNorm	NO	YES	NO
1309799	eplerenone	298869	RxNorm	NO	YES	NO
1346686	eprosartan	83515	RxNorm	NO	YES	NO
1353776	felodipine	4316	RxNorm	NO	YES	NO
1363749	fosinopril	50166	RxNorm	NO	YES	NO
956874	furosemide	4603	RxNorm	NO	YES	NO
1344965	guanfacine	40114	RxNorm	NO	YES	NO
1373928	hydralazine	5470	RxNorm	NO	YES	NO
974166	hydrochlorothiazide	5487	RxNorm	NO	YES	NO
978555	indapamide	5764	RxNorm	NO	YES	NO
1347384	irbesartan	83818	RxNorm	NO	YES	NO
1326012	isradipine	33910	RxNorm	NO	YES	NO
1386957	labetalol	6185	RxNorm	NO	YES	NO
1308216	lisinopril	29046	RxNorm	NO	YES	NO
1367500	losartan	52175	RxNorm	NO	YES	NO
1305447	methyldopa	6876	RxNorm	NO	YES	NO
907013	metolazone	6916	RxNorm	NO	YES	NO
1307046	metoprolol	6918	RxNorm	NO	YES	NO
1309068	minoxidil	6984	RxNorm	NO	YES	NO
1310756	moexipril	30131	RxNorm	NO	YES	NO
1313200	nadolol	7226	RxNorm	NO	YES	NO
1314577	nebivolol	31555	RxNorm	NO	YES	NO
1318137	nicardipine	7396	RxNorm	NO	YES	NO
1318853	nifedipine	7417	RxNorm	NO	YES	NO
1319880	nisoldipine	7435	RxNorm	NO	YES	NO
40226742	olmesartan	321064	RxNorm	NO	YES	NO
1327978	penbutolol	7973	RxNorm	NO	YES	NO
1373225	perindopril	54552	RxNorm	NO	YES	NO
1345858	pindolol	8332	RxNorm	NO	YES	NO
1350489	prazosin	8629	RxNorm	NO	YES	NO
1353766	propranolol	8787	RxNorm	NO	YES	NO
1331235	quinapril	35208	RxNorm	NO	YES	NO

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1334456	ramipril	35296	RxNorm	NO	YES	NO
970250	spironolactone	9997	RxNorm	NO	YES	NO
1317640	telmisartan	73494	RxNorm	NO	YES	NO
1341238	terazosin	37798	RxNorm	NO	YES	NO
942350	torsemide	38413	RxNorm	NO	YES	NO
1342439	trandolapril	38454	RxNorm	NO	YES	NO
904542	triamterene	10763	RxNorm	NO	YES	NO
1308842	valsartan	69749	RxNorm	NO	YES	NO
1307863	verapamil	11170	RxNorm	NO	YES	NO

8.3.6 [QBA eval] THZs

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
1395058	chlorthalidone	2409	RxNorm	NO	YES	NO
974166	hydrochlorothiazide	5487	RxNorm	NO	YES	NO
978555	indapamide	5764	RxNorm	NO	YES	NO
907013	metolazone	6916	RxNorm	NO	YES	NO

---

9 Outcome cohort definition

9.1 Ischemic stroke events during inpatient or emergency room visits

9.1.1 Cohort Entry Events

People may enter the cohort when observing any of the following:

1. condition occurrences of 'Cerebral infarction', starting on or after January 1, 2010.

Restrict entry events to having at least 1 visit occurrence of 'Inpatient or Inpatient/ER visit', starting anytime on or before cohort entry start date and ending between 0 days before and all days after cohort entry start date.

9.1.2 Additional Inclusion Criteria

I. has no events in prior 'clean window' - 365 days

Entry events having no condition occurrences of 'Cerebral infarction', starting in the 365 days prior to cohort entry start date; allow events outside observation period; having at least 1 visit occurrence of 'Inpatient or Inpatient/ER visit', starting anytime on or before

'Cerebral infarction' start date and ending between 0 days before and all days after 'Cerebral infarction' start date.

9.1.3 Cohort Exit

The cohort end date will be offset from index event's start date plus 1 day.

9.1.4 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

9.1.5 Inpatient or Inpatient/ER visit

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
262	Emergency Room and Inpatient Visit	ERIP	Visit	NO	YES	NO
9201	Inpatient Visit	IP	Visit	NO	YES	NO

9.1.6 Cerebral infarction

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
443454	Cerebral infarction	432504007	SNOMED	NO	YES	NO
40479572	Infarct of cerebrum due to iatrogenic cerebrovascular accident	441526008	SNOMED	YES	YES	NO
4046360	Lacunar infarction	230698000	SNOMED	YES	YES	NO
372435	Periventricular leukomalacia	230769007	SNOMED	YES	NO	NO
377254	Multi-infarct dementia, uncomplicated	70936005	SNOMED	YES	NO	NO
379778	Multi-infarct dementia	56267009	SNOMED	YES	NO	NO
443790	Multi-infarct dementia with delusions	25772007	SNOMED	YES	NO	NO
443864	Multi-infarct dementia with depression	14070001	SNOMED	YES	NO	NO

Concept ID	Concept Name	Code	Vocabulary	Excluded	Descendants	Mapped
444091	Multi-infarct dementia with delirium	10349009	SNOMED	YES	NO	NO
4046089	Vascular dementia of acute onset	230285003	SNOMED	YES	NO	NO
4046090	Mixed cortical and subcortical vascular dementia	230287006	SNOMED	YES	NO	NO
4129534	Pituitary apoplexy	237701005	SNOMED	YES	NO	NO

10 Data sources

Data source	Short name	Description
Optum(c) de-identified Electronic Health Record Dataset	optum_ehr	Optum(c) de-identified Electronic Health Record Dataset is derived from dozens of healthcare provider organizations in the United States (that include more than 700 hospitals and 7,000 Clinics treating more than 103 million patients) receiving care in the United States. The medical record data includes clinical information, inclusive of prescriptions as prescribed and administered, lab results, vital signs, body measurements, diagnoses, procedures, and information derived from clinical Notes using Natural Language Processing (NLP).
Optum(c) de-Identified Clinformatics Data Mart Database <96> Date of Death	optum_dod	Optum(c) De-Identified Clinformatics(c) Data Mart Database is an adjudicated administrative health claims database for members with private health insurance, who are fully insured in commercial plans or Medicare Advantage. The population is primarily representative of US commercial claims patients (0-65 years old) with some Medicare (65+ years old) however ages are capped at 90 years. It includes data captured from administrative claims processed from inpatient and outpatient medical services and prescriptions as dispensed, as well as results for outpatient lab tests processed by large national lab vendors who participate in data exchange with Optum. Optum DOD also provides

Data source	Short name	Description
		date of death (month and year only) for members with both medical and pharmacy coverage from the Social Security Death Master File (however after 2011 reporting frequency changed due to changes in reporting requirements) and location information for patients is at the US state level.
IBM MarketScan Commercial Claims and Encounters Database	truven_ccae	IBM MarketScan Commercial Claims and Encounters Database (CCAE) is a US employer-based private-payer administrative claims database. The data include adjudicated health insurance claims (e.g., inpatient, outpatient, and outpatient pharmacy) as well as enrollment data from large employers and health plans who provide private healthcare coverage to employees, their spouses, and dependents. Additionally, it captures laboratory tests for a subset of the covered lives. This administrative claims database includes a variety of fee-for-service, preferred provider organizations, and capitated health plans.
IBM MarketScan MultiState Medicaid Database	truven_mdcd	IBM MarketScan Multi-State Medicaid Database (MDCD) contains adjudicated US health insurance claims for Medicaid enrollees from multiple states and includes hospital discharge diagnoses, outpatient diagnoses and procedures, and outpatient pharmacy claims as well as ethnicity and Medicare eligibility. Members maintain their same identifier even if they leave the system for a brief period; however, the dataset lacks lab data.
IBM MarketScan Medicare Supplemental and Coordination of Benefits Database	truven_mdcr	IBM MarketScan Medicare Supplemental and Coordination of Benefits Database (MDCR) represents health services of retirees in the United States with primary or Medicare supplemental coverage through privately insured fee-for-service, point-of-service, or capitated health plans. These data include adjudicated health insurance claims (e.g., inpatient, outpatient, and outpatient pharmacy). Additionally, it captures laboratory tests for a subset of the covered lives.