Defining a Reference Set for Evaluating the Performance of Active Surveillance Methods

Patrick Ryan
Last revised: 31 January 2010

EXECUTIVE SUMMARY

The OMOP analyses seek to identify methods that properly identify true associations and minimize false positive findings.

OMOP has developed two sets of experiments for evaluating methods, based on the intended use of the analysis:

1. **Monitoring of Health Outcomes of Interest (HOIs):** to provide a reliable estimate the strength of the association between exposure and a defined outcome.

    - A set of 10 HOIs have been defined that are of particular interest to stakeholders due to their medical interest, their relatively clear relationship to drug toxicity, and their potential public health impact and their association to a particular drug exposure in clinical trials or reported from other observational studies confirmed, are the test cases for the **Monitoring of HOIs**.

    - A reference set of 'negative controls' of drug-condition pairs where no temporal association should be expected. These negative controls will serve the basis of assessing the degree to which methods generate false positive findings. Eligible pairs were identified from the drug-HOI test cases if the outcome was not listed on the product label for the drug as either a possible adverse event or related to the drug's indication.

2. **Identification of non-specified associations:** exploratory analysis aims to generate hypotheses from observational data by identifying associations between exposure and conditions for which the relationships were previously unknown.
    - Non-specified associations will be derived using individual data components as markers for potential relationships: conditions will be identified from diagnosis codes recorded on administrative claims or as part of an electronic health record problem list.
    - OMOP experiments will focus exclusively on method's ability to identify relationships between drug exposure and condition occurrence,
    - A systematic analysis (SPLICER software) of the structured product labels (SPL) of the drugs used in the drug-HOI test cases provided a reference set of 'true associations' and 'negative controls' using labeled events as surrogate markers for 'true associations' and selecting terms unrelated to any labeled events as 'negative controls'.

- Two levels of classification were used: Tier 1 events are those conditions that occur in either the 'Black box' or 'Warnings/Precautions' sections on >=50% of the SPLs within the DOI. Tier 2 events are those conditions that occur as adverse events anywhere on the product label (Black box, warnings/precautions, adverse reactions, or Post-marketing experience) on >=50% of the SPLs for a given exposure.
- It could be argued that events listed in black box warnings or warnings/precautions are more likely to be causally related and observable. Primary analyses for OMOP will be based on all Tier 2 events, but Tier 1 classification offers a potential sensitivity analysis when assessing methods performance.

- **Sufficient sample size of conditions to test of methods** is examined in prevalence rates of conditions in database of 59 million persons. Both the 'true associations' and 'negative controls' include at least 20 conditions that occur in >1% of the population, as well as over 50 conditions that occur <0.1% of persons. The 43 Tier 1 warning events are generally less frequent, but still provide some diversity in observed prevalence rates.

- **Potential limitations**: Because 'truth' is not known for any drug, we are required to select some surrogate (which has its own undefined sensitivity and specificity).
  - We understand that labeled events have not necessarily been shown to be causally related to drug, or may not be expected to be observed in subsequent study. In particular, adverse events listed in the Adverse Reactions and Post-Marketing Experience section may reflect occurrence from clinical trials or spontaneous reporting without any expectation of causality.
  - It is possible that 'negative controls' have been selected that do have legitimate temporal relationships with the drugs of interest, and either have not been previously identified or were not listed on the product label.
  - The number of test cases can be considered the sample size within this methodological experiment. Because the same set of test cases is being consistently applied across all methods, any misclassification of test cases (either 'true associations' that are not related, or 'negative controls' that have an association) should not introduce differential bias in the experiment and should not influence the relative assessment of Mean Average Precision scores between methods.

Background

A national active surveillance system will involve a systematic process for analyzing multiple observational healthcare data sources, including administrative claims and electronic health records, to better understand the effects of medical products by estimating temporal relationships between exposure and outcomes. The active surveillance system can be used to 1) characterize known side effects, 2) monitor preventable adverse events, and 3) explore remaining uncertainties. The goal of the active surveillance system is to contribute supplemental information to other existing sources of safety information (including pre-clinical data, clinical trials, and spontaneous adverse event reporting) to support decision-making about appropriate use of medical products for regulatory agencies, providers, and patients.

The Observational Medical Outcomes Partnership (OMOP, http://omop.fnih.org) is a public-private partnership conducting methodological research to inform the appropriate use of observational data for active surveillance. A key component of this initiative is to empirically evaluate the performance of alternative analysis approaches in their ability to identify known associations and discern from false positive findings.

OMOP has developed two sets of experiments for evaluating methods, based on the intended use of the analysis.

- **Monitoring of Health Outcomes of Interest:** The goal of this surveillance analysis is to monitor the relationship between a series of drugs and specific outcomes of interest. These analyses require an effective definition of the events of interest in the context of the available data (e.g. 'acute liver injury' may best be defined by a combination of medical diagnoses, pharmacy records, procedure codes, and/or laboratory results). The goal of these analyses is to provide a reliable estimate the strength of the association between exposure and outcome.

- **Identification of non-specified associations:** This exploratory analysis aims to generate hypotheses from observational data by identifying associations between drugs and conditions for which the relationships were previously unknown. This type of analysis is likely to be considered an initial step of a triaged review process, where many drug-outcome pairs are simultaneously explored to prioritize the drugs and outcomes that warrant further attention. Unlike 'Monitoring of Health Outcomes of Interest', where analyses utilize pre-specified outcome definitions, 'Identification of non-specified associations' uses available data, such as condition occurrence, as surrogate markers for outcomes to facilitate this initial exploratory analysis. In this regard, information generated from these analysis should inform, but is unlikely to be sufficient, for the development of a more refined hypothesis about a specific drug-outcome relationship. It could be expected that a primary consideration for identification analyses is developing an efficient model to allow high-throughput computing across large sets of potential drug-outcome relationships.

For each analysis purpose above, there are two aims: 1) to measure the degree to which methods properly identify true associations and minimize false positive findings when applied across an entire database, and 2) to measure the performance of methods as data accumulates over time. Conceptually, there is no reason why each method cannot be used for both purposes, as the analyses only differ in the definition of the outcome. However, there may be practical considerations for why some methods are not fully scaleable for assessing many drug-condition pairs for 'identification of non-specified associations', so may only be tested against a more restricted set of test cases for 'monitoring of Health Outcomes of Interest'.

To facilitate these assessments, a reference set is required that provides a pre-defined benchmark to base comparisons of methods against. In this context, the reference set is a list of test cases of drug-condition pairs classified dichotomously as 'true associations' or 'negative controls'. Methods will then be executed against these test cases, and the estimates derived from the methods will be compared to the benchmark. For OMOP, mean average precision (MAP) has been selected as the objective measure of performance. In short, estimates are rank-ordered by score, and precision is estimated at each possible threshold (number of 'true associations' above threshold / total number of outcomes above threshold), before an average is taken across these precision estimates. Mean average precision is also being used to measure performance of methods against the simulated data as part of the OMOP Cup and is described at length for those purposes (http://omopcup.orwik.com). This document describes the OMOP reference set used for each analysis program and the rationale that was used to develop them.

Monitoring of Health Outcomes of Interest

While all health outcomes may be considered within observational pharmacovigilance, certain health outcomes are of particular interest to regulatory agencies, industry, providers and patients due to their medical interest, their relatively clear relationship to drug toxicity, and their potential public health impact. Health outcomes of interest (HOI) are a subset of all conditions which are of particular focus due to their historical associations and toxicities with drugs, medical significance, and/or public health implications. Many health outcomes of interest are potential safety concerns that could be regularly monitored in observational data, including conditions historically identified as 'designated medical events'. Additionally, specific health service utilization indicators (e.g. number of hospitalizations) and the occurrence of specific abnormal laboratory values (e.g. Hy's law) may be considered as HOIs. However, health outcomes of interest could also include potential beneficial outcomes.

It is important to note that there are some HOIs that will be actively monitored for all drugs, but individual drugs may have additional specific HOIs that require surveillance due to hypotheses generated from other sources prior to approval. Within OMOP, the primary distinction between HOIs and non-specified conditions is the need for a thorough review of the outcome definition, but it is anticipated that the HOI definition process developed should be useful for supporting active surveillance of all types of HOIs moving forward.

A systematic process was followed to identify the preliminary list of health outcomes of interest under consideration within OMOP.

1. The potential list of HOIs of interest were initially generated from 4 sources:
    a. The 'Always Expedited List' (21 CFR Parts 310, 312, et al. Safety Reporting Requirements for Human Drug and Biological Products)
    b. The adverse events responsible for market withdrawal
    c. The adverse events responsible for black box warnings
    d. Additional suggestions from OMOP participants
2. OMOP participants were then asked to 'rank' order this list on 3 levels: 'Imperative', 'important', and 'nice to have.' This focused list was used as the starting point for identifying the drug-HOI pairs to be used in subsequent proposals.
3. This ranking was combined with an assessment of the HOI appearing on the Always Expedited list, the market withdrawal list, and the black box warning list to assure that the most impactful HOIs were considered for selection. The list was then reviewed and additions or deletions made.

Table 1 highlights the initial pool of potential health outcomes of interest.

Table 1: OMOP Initial Health Outcome of Interest List

OMOP Health Outcomes of Interest List

	Source		
	Always expedited / DMEs / others	**Triggers for withdrawals**	**Triggers for black box**
Safety			
Acute liver failure	***X***	***X***	***X***
Acute pancreatitis	X		
Acute respiratory failure	X		
Aggressive and/or violent behaviors			X
Agranulocytosis	X		
Anaphylaxis	X		X
Angioedema	X		
Aplastic anemia	X		
Autoimmune disorders			X
Birth defects		X	
Blindness	X		
Brain infection		X	
Cardiopulmonary arrest		X	
Congenital anomalies	X		
Death	X		
Deep vein thrombosis			X
Diffuse intravascular coagulopathy	X		
Endometrial cancer			X
Endotoxin shock	X		
Fatal allergic reaction		X	
Fatal arrhythmia		X	
Fatal infections and bleeding			X
Fractures			
GI perforation	X		X
Heart valve disease		X	
Hemolytic anemia	X	X	
Increased risk of cardiovascular events			X
Increased risk of suicidality			X
Infectious disorders			X
Intrathecal baclofen withdrawal leading to life threatening sequelae and/or death			X
Invasive breast cancer			X
Ischemic colitis		X	
Ischemic disorders			X
Life-threatening asthma			X
Life-threatening diarrhea			X
Life-threatening skin reactions			X
Liver necrosis	X		
Loss of consciousness			X
Loss of significant bone mineral density			X
Malignant hypertension	X		
Muscle damage leading to kidney failure		X	
Myocardial infarction	***X***	***X***	***X***
Myocarditis			X
Nephrogenic systemic fibrosis			X
Neuropsychiatric disorders			X

OMOP Health Outcomes of Interest List (*continued*)

	Source		
	Always expedited / DMEs / others	**Triggers for withdrawals**	**Triggers for black box**
Possible risk of cancer			X
Pulmonary embolism			X
Pulmonary fibrosis	X		
Pulmonary hypertension	X		
Renal failure	X	X	
Rhabdomyolysis	X		
Sclerosing syndromes	X		
Secondary acute myelogenous leukemia			X
Seizures	X		
Serious blood clots		X	
Serious drug-drug interactions leading to muscle damage and fatal arrhythmia		X	
Severe breathing difficulty		X	
Severe constipation		X	
Severe cutaneous or mucocutaneous reaction			X
Skin disease		X	
Stevens-Johnson syndrome	X		
Stroke	***X***	***X***	***X***
Thrombotic thrombocytopenic purpura	X		
Torsades de Pointe	X		
Toxic epidermal necrolysis	X		
Transmission of an infectious agent by marketed drug or biologic	X		
Uterine sarcoma			X
Ventricular fibrillation	X		
Visual disturbances			X
Benefits Suggested by Workgroup Members			
Clinical outcomes: changes in frequency or severity of clinical endpoints of relevance			
Costs associated with services provided			
Patient persistence / compliance			
Productivity (absenteeism)			
Quality of life			
Utilization of healthcare services			

Once the initial pool of HOIs was identified, the objective was to identify test cases of known associations between drugs and those outcomes. It was felt that the pilot should focus on replicating drug-HOI associations that had previously been reported in the literature using observational data. Too often, drug-HOI relationships have been reported or have arisen from other sources (spontaneous AE data, clinical series) but were not investigated or confirmed in observational data. The requirement of prior assessment using observational data is intended to assure that other investigators had determined a way to define both the drug and the condition(s) of interest and the pilot work would then be confirmatory; in other words, the OMOP would have a higher likelihood of finding the association and if not, further consideration would need to be given to the approach.

The process and requirements to identify candidate HOI-drug pairs was as follows:

1. From the focused list of HOIs, a systematic literature review was undertaken to attempt to identify reported findings from observational studies in claims or EHR data that involve specific drugs associated with the HOI.
2. Outcomes identified in observational studies were confirmed to be reported in the product label of at least one drug in the class.
3. Drug-condition pairs were excluded that involve products under patent protection to avoid complicating the pilot with the associated Regulatory requirements for reporting; this will result in the selection of more mature products for study.
4. The Principal Investigators selected a restricted set of drug-condition pairs from the candidate list.

Table 2 provides the details about the specific drug-HOI pairs selected as 'true associations' for OMOP's research purposes.

There are clearly limitations of drug-HOI selections. While the goal is to select drug-HOI pairs to be representative of the types of project expected in the future, it may not be a sufficient sample to draw conclusions for all scenarios. The requirement of a prior report in the literature from another observational database will substantially restrict our sample, as will the requirement for non-proprietary (and therefore generic) drugs or classes. While the association has been reported in a product label and observed in a prior study, it is possible any of the pairs selected do not reflect true causal relationships and should not be expected to be observed. Moreover, it is possible that the association between the drug and the HOI will not be observed in observational data because of changes in clinical practice to mitigate known risks once it becomes common knowledge. This concern is particularly relevant to associations that are known for a long time or have garnered the most attention. Because of our selection criteria requiring the drugs to be mature, it is likely that association has been known for a long time.

Finally, because of the rules implemented for HOI-drug pairs (mature drug and literature-reported analysis in observational data), it is difficult to test the use of observational data in 'newly launched' product scenarios; the examples use mature products and the population of users later in the lifecycle of these products may not represent the earlier users. Our proposal will attempt to reconstruct and approximate newly initiated drug use within the study design by selecting incident users, however it will likely reflect user characteristics of the population after the product has been marketed (and prescribing physicians have experience with the drug) and not necessarily the population of users at, or soon after, product launch. It is also important to remember that observational data, particularly claims, may not have exposure to a product for at least a year after approval as it takes time for drugs to be adopted onto formularies.

Selecting negative controls

After the 10 Health Outcomes of Interest (HOI) and 10 drugs of interest (DOI) were identified, the next step was to define a reference set of 'negative controls' of drug-condition pairs where no temporal association should be expected. These negative controls will serve the basis of assessing the degree to which methods generate false positive findings. The decision was made

to use the same HOI and same DOI, but to identify different drug-outcome combinations for which no evidence. Negative control pairs were identified if the outcome was not listed on the product label for the drug of interest as either a possible adverse event or related to the drug's indication.

Because the HOIs of 'Bleeding' and 'GI Ulcer hospitalization' are correlated, drugs not related to either condition were selected for only one negative control test case. Similarly, negative controls for drugs were selected for either 'Myocardial Infarction' or 'Mortality after MI', but not both. No negative controls were selected for the outcome 'Hospitalization' because it all drug indications are for serious conditions that could result in hospitalization, and it should not be assumed that an increase in temporal relationship between exposure and hospitalization should be attributed to the drug.

Table 3 highlights the HOI-DOI matrix with the 'true associations' and 'negative controls'. Each method will be applied to 100 test cases (10 HOI x 10 DOI), but the MAP score will be calculated using on the estimates for the 9 safety 'true associations' and 44 'negative controls'. The two 'true positive benefit' associations will be analyzed separately.

It is important to recognize the objective of the OMOP experiment is to assess the performance of various analysis methods for active surveillance, and not to generate clinically relevant information about specific drug-condition associations. To that end, the precision in classification of 'true associations' and 'negative controls' may be imperfect but sufficient for experimentation, because the full set of test cases will be consistently applied across all methods. The mean average precision metric should not be seen as a definitive predictive measure of a methods' future expected performance, but instead as a relative barometer to compare methods under specific circumstances.

Table 2: Drug-HOI pairs uses as 'true associations' for OMOP experiments

Drug/class: product names	Health Outcome of Interest	Outcome Type: 1- Benefit or 2- Safety	Product Maturity: 1- all generic, 2- some, 3- all patented	Outcome rate: 1- Low or 2- high background rate	Time to onset: 1- acute, 2- insidious, 3- delayed	Journal References (specific product mentions in red)	Labeled
ACE inhibitors: Captopril, Enalapril, Lisinopril	Angioedema	Safety	2	1	1	(1-3) Miller: all ACE (mention lisinopril, fosinipril, captopril) Brown 1996: restricted to captopril, enalapril, lisinopril; Brown 1997: ACE (no mention of specific products)	yes: angioedema
ACE inhibitors: benazepril, captopril, enalapril, fosinopril, lisinopril, quinapril, and ramipril.	Hospitalization (including readmission and mortality)	Benefit	2	2	2	(4-10) Luthi 2003: benazepril, captopril, enalapril, fosinopril, lisinopril, quinapril, and ramipril; Ahmed 2003: ACE (no mention of specific products); Trewet 2007: enalapril, benezepril, captopril, fosinopril, lisinopril, moexipril, quinapril, ramipril, perindopril	Yes
Amphotericin B	Renal failure	Safety	1	1	2	(11, 12) Bates 2001: parenteral amphotericin B	yes decreased renal function
Antibiotics: erythromycins, sulfonamides,	Acute liver injury (symptomatic	Safety	1	1	2	(13-15) Carson 1993: Erythromycins (mention ethylsuccinate),	yes multiple hepatic-

Drug/class: product names	Health Outcome of Interest	Outcome Type: 1- Benefit or 2- Safety	Product Maturity: 1- all generic, 2- some, 3- all patented	Outcome rate: 1- Low or 2- high background rate	Time to onset: 1- acute, 2- insidious, 3- delayed	Journal References (specific product mentions in red)	Labeled
and tetracyclines	hepatitis)					Sulfonamides, tetracyclines (mention tetracycline hydrochloride, doxycycline, minocycline); Heaton 2007: tetracycline, doxycycline;	AE types
Antiepileptics: carbamazepine, valproic acid, and phenytoin	Aplastic anemia	Safety	1	1	2	(18) Handoko 2006: carbamazepine, valproic acid, phenytoin	yes aplastic anemia, bone marrow supression, pancytopenia
Benzodiazepines	Hip fracture	Safety	1	2	2	(19-22) Wang 2001: benzodiazepines (no mention of specific products); Wang AJP 2001: alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, estazolam, flurazepam, halazepam, lorazepam, oxazepam, prazepam, quazepam, temazepam, or triazolam; Ray 1989: chlordiazepoxide, clorazepate, diazepam, flurazepam, alprazolam, bromazepam, lorazepam, oxazepam, triazolam	Yes

Drug/class: product names	Health Outcome of Interest	Outcome Type: 1- Benefit or 2- Safety	Product Maturity: 1- all generic, 2- some, 3- all patented	Outcome rate: 1- Low or 2- high background rate	Time to onset: 1- acute, 2- insidious, 3- delayed	Journal References (specific product mentions in red)	Labeled
Beta blockers	Mortality after MI	Benefit	1	2	2	(23-29) Krumholz 1999: beta blockers (no mention of specific products); Barron 1998: propanolol, metoprolol, atenolol	Yes
Bisphosphonates: alendronate only	GI ulcer hospitalizations	Safety	2	2	2	(35-37) Ettinger 1998: alendronate; Donahue 2002: alendronate; Halkin 2007: alendronate	Yes
Tricyclic antidepressants	Myocardial infarction	Safety	1	2	2	(38-40) Tata 2005: Tricyclic antidepressants (mention dothiepin, amitriptyline, lofepramine); Cohen 2000: tricylic agents (all listed in AHFS codes)	Yes MI, arrhythmias
Typical antipsychotics	Myocardial infarction	Safety	1	2	2	(41-43) Enger 2004: Chlorpromazine, Fluphenazine decanoate, Fluphenazine enanthate, Fluphenazine hcl, Haloperidol, Haloperidol decanoate, Loxapine hcl, Loxapine succinate, Mesoridazine,Molindone, Perphenazine and amitriptyline hcl/perphenazine, Pimozide,Prochlorperazine edisylate, Prochlorperazine maleate,	Clozaril: CHF, myocarditis

Drug/class: product names	Health Outcome of Interest	Outcome Type: 1- Benefit or 2- Safety	Product Maturity: 1- all generic, 2- some, 3- all patented	Outcome rate: 1- Low or 2- high background rate	Time to onset: 1- acute, 2- insidious, 3- delayed	Journal References (specific product mentions in red)	Labeled
						Promazine, Propiomazine, Thioridazine, Thiothixene, Trifluoperazine, Triflupromazine; Nakagawa 2006: flupentixol, fluphenazine, haloperidol, pimozide, periciazine, perphenazine, prochlorperazine, zuclopenthixol, chlorprothixene, chlorpromazine, levomepromazine, melperon, pipamperone, promazine and thioridazine	
Warfarin	Bleeding	Safety	1	2	2	(44-48) Zhang 2006: Warfarin (all doses); Hollowell 2003: warfarin (no mention of specific products)	Yes

Table 3: Test cases for 'Monitoring of Health Outcomes of Interest'

Test cases to be used for evaluating method performance for 'Monitoring of Health Outcomes of Interest'

	Outcome									
Drug	1. Angioedema	2. Aplastic Anemia	3. Acute Liver Injury	4. Bleeding	5. GI Ulcer Hospital-ization	6. Hip Fracture	7. Hospital-ization	8. Myocardial Infarction	9. Mortality after MI	10. Renal Failure
1. ACE Inhibitors	R	N			N	N	B			
2. Amphotericin B	N	N	N			N			N	R
3. Antibiotics		N	R	N		N		N		N
4. Antiepileptics	N	R			N				N	N
5. Benzodiazapines	N	N	N	N		R		N		N
6. Beta blockers	N	N	N		N	N			B	N
7. Bisphosphonates		N	N		R			N		N
8. Tricyclic antidepressants		N	N	N				R		N
9. Typical antipsychotics					N			R		N
10. Warfarin	N	N		R		N			N	N

Legend	Total
B- 'True positive' benefit	2
R- 'True positive' risk	9
N- 'Negative control'	44
Avoid selection due to labeling	
Not selected due to correlation with HOI	

Identification of Non-Specified Associations

Identification of non-specified associations requires estimation of temporal relationships between exposure and outcome, without necessarily the hindsight needed to pre-specify the outcome definition. As such, methods make use of data elements already available to identify possible relationships with drug use. When viewed in an exploratory framework, these relationships may inform and support developing a refined hypothesis about a potential drug-related clinical effect for further study, but each relationship in isolation will generally be regarded as insufficient to draw reliable inferences about a causal relationship between exposure and outcome.

Whereas a Health Outcome of Interest may be defined by any combination of multiple diagnosis codes, potentially preceded by diagnostic procedures or laboratory tests and/or followed by treatment procedures or medical use, non-specified associations will be derived using individual data components as markers for potential relationships. The most readily available data element is condition occurrence, such as diagnosis codes recorded on administrative claims or as part of an electronic health record problem list. OMOP experiments will focus exclusively on method's ability to identify relationships between drug exposure and condition occurrence, though it should be recognized that methods could have further utility in exploring other temporal relationships with procedure occurrence, other drug exposure, health service utilization, and other clinical observations as well.

For purposes of assessment of method performance for 'identification of non-specified conditions', OMOP will use the same 10 drugs of interest (DOI), as selected for 'monitoring of Health Outcomes of Interest'. These DOIs are:

- ACE inhibitors
- Amphotericin B
- Antibiotics (erythromycin, sulfonamides, tetracyclines)
- Antiepileptics (carbamazapine, phenytoin)
- Warfarin
- Benzodiazepines
- Bisphosphonates: Alendronate
- Tricyclic antidepressants
- Typical antipsychotics
- Beta blockers

The specific medical products that comprise each DOI are described elsewhere, and were defined largely based on the prior literature used to select the drug-HOI pairs. As such, drug classes such as ACE inhibitors and Beta blockers contain more products than the drugs selected to comprise the DOI definition.

For each of the 10 DOIs, the objective is to define a reference set of 'true associations' and 'negative controls' that can be used to assess methods' performance. This objective was achieved through a systematic analysis of the structured product labels within each DOI, using

labeled events as surrogate markers for 'true associations' and selecting terms unrelated to any labeled events as 'negative controls'.

Regenstrief Institute has developed a novel application, SPLICER, which performs natural language processing on structured product labels (SPL) to extract terms that may be adverse events. The application classifies the events by the location of occurrence, as 'Black box', 'Warnings and Precautions', 'Adverse Reactions' or 'Post-marketing experience', and codes the terms of MedDRA preferred terms (PTs). Each SPL was mapped to a corresponding RxNorm drug concept.

SPLICER's most recent run was performed on 12/19/2009 and include 5602 SPL's from the DailyMed site. This set of labels comprised 1706 distinct generic drugs and 2861 distinct brand names. SPLICER successfully coded and extracted 608,948 adverse events from these labels. These events were mapped to 4627 distinct MedDRA preferred terms. An evaluation of SPLICER's performance in retrieving events from the Adverse Reaction section of 100 labels demonstrated a recall of 93% and a precision of 95%. The output of SPLICER for SPLs from the OMOP drugs of interest was used for defining the reference set.

Table 4 summarizes the OMOP drugs of interest, including the number of distinct SPLs used for in the analysis and the number of unique ingredients reflected within those SPLs. The ACE inhibitor definition includes 9 ingredients, represented by 119 different product labels.

Table 4: OMOP Drugs of Interest summary

DOI_NAME	DOI_DESCRIPTION	LABEL COUNT	INGREDIENT COUNT
OMOP ACE Inhibitor	ACE inhibitors: benazepril, captopril, enalapril, fosinopril, lisinopril, quinapril, and ramipril; restricted to oral form	119	9
OMOP Amphotericin B	parenteral Amphotericin B	4	1
OMOP Antibiotics: erythromycins, sulfonamides, and tetracyclines	Antibiotics: erythromycins, sulfonamides, and tetracyclines; restricted to oral and injectable	49	6
OMOP Antiepileptics: carbamazepine, phenytoin	Antiepileptics: carbamazepine, phenytoin; restricted to oral and injectable	25	3
OMOP Benzodiazepines	Benzodiazepines: alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, estazolam, flurazepam, halazepam, lorazepam, oxazepam, prazepam, quazepam, temazepam, or triazolam	97	10
OMOP Beta blockers	Beta blockers: propanolol, metoprolol, atenolol; restricted to oral form	53	3
OMOP Bisphosphonates	Bisphosphonates: alendronate	8	1
OMOP Tricyclic antidepressants	Tricyclic antidepressants: restricted to oral and injectable	35	7

DOI_NAME	DOI_DESCRIPTION	LABEL COUNT	INGREDIENT COUNT
OMOP Typical antipsychotics	Typical antipsychotics: Chlorpromazine, chlorprothixene, levomepromazine, flupentixol, Fluphenazine decanoate, Fluphenazine enanthate, Fluphenazine hcl, Haloperidol, Haloperidol decanoate, Loxapine hcl, Loxapine succinate, melperon, Mesoridazine, Molindone, Perphenazine, amitriptyline hcl/perphenazine, Pimozide, pipamperone, promazine, Prochlorperazine edisylate, periciazine, Prochlorperazine maleate, Promazine, Propiomazine, Thioridazine, Thiothixene, Trifluoperazine, zuclopenthixol	41	10
OMOP Warfarin	Warfarin	6	1

Table 5 provides a descriptive summary of the number of events extracted from each SPL, summarized by the ingredients within each DOI. For example, lisinopril is one of nine ingredients within the ACE Inhibitor group; there are 28 distinct SPLs that involve lisinopril. Among those 28 labels, SPLICER identified 234 distinct MedDRA PTs. On average, each of the labels listed 184 distinct events, with the minimum of 60 events and a maximum of 205 events. This table highlights the variability observed in product labeling, both among labels for the same ingredient as well as among ingredient within the OMOP DOI groupings.

Table 5: SPL event summary by ingredient

OMOP Drug of Interest	Ingredient name	Number of SPLs	Distinct events across SPLs	Min events among SPLs	Average events among SPLs	Max events among SPLs
OMOP ACE Inhibitor	Lisinopril	28	234	60	184	205
	moexipril	6	261	72	158	242
	quinapril	15	174	72	101	151
	Ramipril	9	112	87	100	110
	benazepril	19	180	58	89	133
	Captopril	12	143	103	114	135
	Enalapril	15	211	117	142	171
	Fosinopril	12	183	116	132	140
	Perindopril	3	153	150	151	152
OMOP Amphotericin B	Amphotericin B	4	260	84	126	171
OMOP Antibiotics	Oxytetracycline	2	45	23	27	30
	Sulfanilamide	1	3	3	3	3
	Erythromycin	31	114	2	15	94
	Sulfadiazine	1	55	55	55	55
	Sulfamethoxazole	11	94	73	79	86
	Tetracycline	3	136	40	72	130
OMOP Antiepileptics	Mephenytoin	1	56	56	56	56
	Carbamazepine	12	156	106	115	135

OMOP Drug of Interest	Ingredient name	Number of SPLs	Distinct events across SPLs	Min events among SPLs	Average events among SPLs	Max events among SPLs
	Phenytoin	12	67	47	49	55
OMOP Benzodiazepines	Triazolam	4	78	77	77	78
	Diazepam	15	111	29	45	71
	Oxazepam	4	37	34	36	37
	Estazolam	5	117	113	115	116
	Flurazepam	2	57	56	57	57
	Alprazolam	21	213	83	129	164
	Lorazepam	20	130	17	70	90
	Clonazepam	8	212	163	192	203
	Temazepam	11	39	21	36	39
	Chlordiazepoxide	7	87	16	44	81
OMOP Beta blockers	Metoprolol	19	156	62	77	112
	Atenolol	17	106	67	71	82
	Propranolol	17	98	49	63	86
OMOP Bisphosphonates	Alendronate	8	72	59	68	70
OMOP Tricyclic antidepressants	Trimipramine	1	87	87	87	87
	Amitriptyline	9	166	79	98	138
	Desipramine	3	103	99	100	101
	Nortriptyline	8	97	86	90	94
	Doxepin	6	59	56	56	57
	Protriptyline	3	113	93	106	113
	Clomipramine	5	329	323	324	326
	Thioridazine	2	92	92	92	92
	Thiothixene	3	86	83	84	86
	Trifluoperazine	2	117	116	116	116
	Molindone	1	28	28	28	28
	Perphenazine	4	196	135	137	138
	Prochlorperazine	6	125	110	115	120
	Fluphenazine	7	107	98	101	104
	Haloperidol	9	91	74	80	84
	Loxapine	4	75	72	73	75
	Chlorpromazine	3	109	103	105	107
OMOP Warfarin	Warfarin	6	61	52	57	59

Selecting 'true associations':

Labeled events were selected as 'true association' test cases if three criteria were satisfied:

1. MedDRA PT was listed on >=50% of structured product labels within the OMOP DOI
2. MedDRA PT had at least one ICD-9-CM code directly mapped to it within the OMOP standardized terminology

3. One of the ICD-9-CM codes mapped to the MedDRA PT also directly mapped to at least one SNOMED concept

We created two levels of classification: Tier 1 events are those conditions that occur in either the 'Black box' or 'Warnings/Precautions' sections on >=50% of the SPLs within the DOI. Tier 2 events are those conditions that occur as adverse events anywhere on the product label (Black box, warnings/precautions, adverse reactions, or Post-marketing experience) on >=50% of the SPLs within the DOI. Tier 1 events are a subset of the Tier 2 labeled events. It could be argued that events listed in black box warnings or warnings/precautions are more likely to be causally related and observable. Primary analyses for OMOP will be based on all Tier 2 events, but Tier 1 classification offers a potential sensitivity analysis when assessing methods performance. Table 6 shows the number of events that satisfy the selection criteria for each DOI. In total, 324 'true association' test cases were identified, of which 43 were Tier 1 events.

Table 6: Test cases by OMOP drugs of interest

	Test case events for 'Identification of non-specified conditions'		
DOI_NAME	Tier 2 LABEL	Tier 1 WARNING	NEGATIVE CONTROL
OMOP ACE Inhibitor	38	12	1253
OMOP Amphotericin B	70	4	1253
OMOP Antibiotics	9	0	1253
OMOP Antiepileptics	19	1	1253
OMOP Benzodiazepines	17	0	1253
OMOP Beta blockers	22	2	1253
OMOP Bisphosphonates	27	9	1253
OMOP Tricyclic antidepressants	32	6	1253
OMOP Typical antipsychotics	28	4	1253
OMOP Warfarin	19	5	1253
Totals	281	43	12530
Total 'true associations' (Tier 1 + 2)		324	
Total test cases			12854

The rationale for criteria #1 was that the majority of labels contributing to the DOI needed to list the event in order to have some confidence that the association could be potentially observed. 50% was identified as a threshold after review of alternative thresholds, as shown in Figure 1. If 'true associations' were defined as events occurring on all SPLs within a DOI, only 68 test cases would have been identified and 3 of the DOIs would have no 'true' events. In contrast, had the threshold been set by selecting any event that occurred on at least 1 SPL, 1110 terms would have been selected but this would have included many events that only existed on one label (and therefore may be considered less likely to be observed when exploring associations across the DOI).

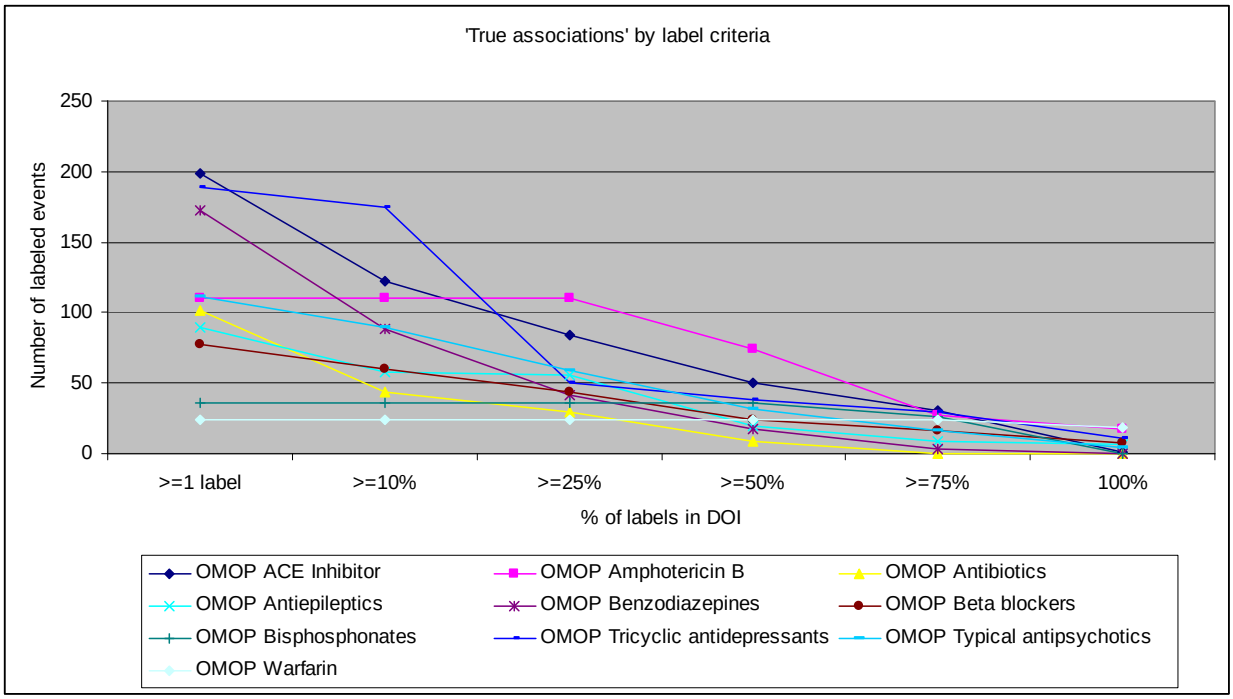

Figure 1: Number of 'true associations' per DOI by label criteria

The rationale for criteria #2 and #3 is to ensure the event is theoretically observable across the OMOP data community. That is, some MedDRA preferred terms include adverse event concepts that have no corresponding codes in ICD-9-CM, so could not possibly be recorded in any US administrative claims system. Another issue is that some ICD-9 codes may map to multiple concepts; in these cases, the ICD-9 code is mapped in the standardized terminology to a surrogate concept and is excluded from consideration as a test case. We chose to restrict our focus to concepts that also have corresponding SNOMED concepts to enable us to evaluate both MedDRA and SNOMED as alternative standardized terminologies for active surveillance.

Selecting 'negative controls':

Labeled events were selected as 'negative control' test cases if three criteria were satisfied:

1. MedDRA PT does not have the same High Level Term as any PT that was extracted from any location (black box, warnings/precautions, adverse reactions, post-marketing experience, indications) of any structured product labels among any of the OMOP DOI
2. MedDRA PT had at least one ICD-9-CM code directly mapped to it within the OMOP standardized terminology
3. One of the ICD-9-CM codes mapped to the MedDRA PT also directly mapped to at least one SNOMED concept
4. MedDRA PT is not among the terms used to define Health Outcomes of Interest
5. MedDRA PT belongs to a System Organ Class other than "Pregnancy, puerperium and perinatal conditions" and "Congenital, familial and genetic disorders"

Criteria #1 ensures that no 'negative control' is related to any labeled event. This is a conservative restriction to avoid selecting any terms that could be drug-related by eliminating all adverse events that occurred on at least one label. The 'negative control' must exist in a High Level Term without any other labeled events to minimize the chance that a 'negative control' would be selected because it was a distinct term even though it was clinically similar. For example, 'myocardial infarction' is a labeled event, but 'acute myocardial infarction' and 'myocardial ischemia' are not; however, since all three terms belong to the HTL 'Coronary ischemic disorders', all are excluded as candidate 'negative controls'. Criteria #4 was applied because the HOI concepts are being treated separately as part of the experiments for 'Monitoring of Health Outcomes of Interest'. Criteria #5 was applied because pregnancy-related adverse events are not the specific focus of OMOP and should be considered separately for future research.

Based on these criteria, 1253 distinct terms were identified as 'negative controls', each of which will be applied across all 10 DOIs to provide 12,530 test cases for methods assessment.

Discussion:

For 'identification of non-specified association' experiments, each method will be executed for 10 drugs of interest and 1,397 distinct conditions. Only the 12,854 'true associations' and 'negative controls' will be used for estimating the Mean Average Precision of each method. The full reference set listing is provided in Appendix 1.

By selecting multiple drugs that representing varying degrees of exposure for different diseased populations, we hope to provide information that can be generalized to some extent as to the utility of the methods in supporting assessment of other drugs. Similarly, by selecting labeled events, it is anticipated that we will have a broad set of conditions to explore as well, ranging in severity, incidence and strength of association, which may inform thoughts about method's performance for other conditions.

One consideration for determining whether the reference set of 'true associations' and 'negative controls' is a sufficient test of methods is that the test cases consist of conditions of various prevalence rates. Figure 2 shows the prevalence within the Thomson Reuters MarketScan commercial claims database (CCAE), an administrative claims database of 59m persons. Both the 'true associations' and 'negative controls' include at least 20 conditions that occur in >1% of the population, as well as over 50 conditions that occur <0.1% of persons. The 43 Tier 1 warning events are generally less frequent, but still provide some diversity in observed prevalence rates. This demonstrates that the reference set should allow for subsequent stratified analyses by observed prevalence to determine if occurrence rates impact ability to detect associations.

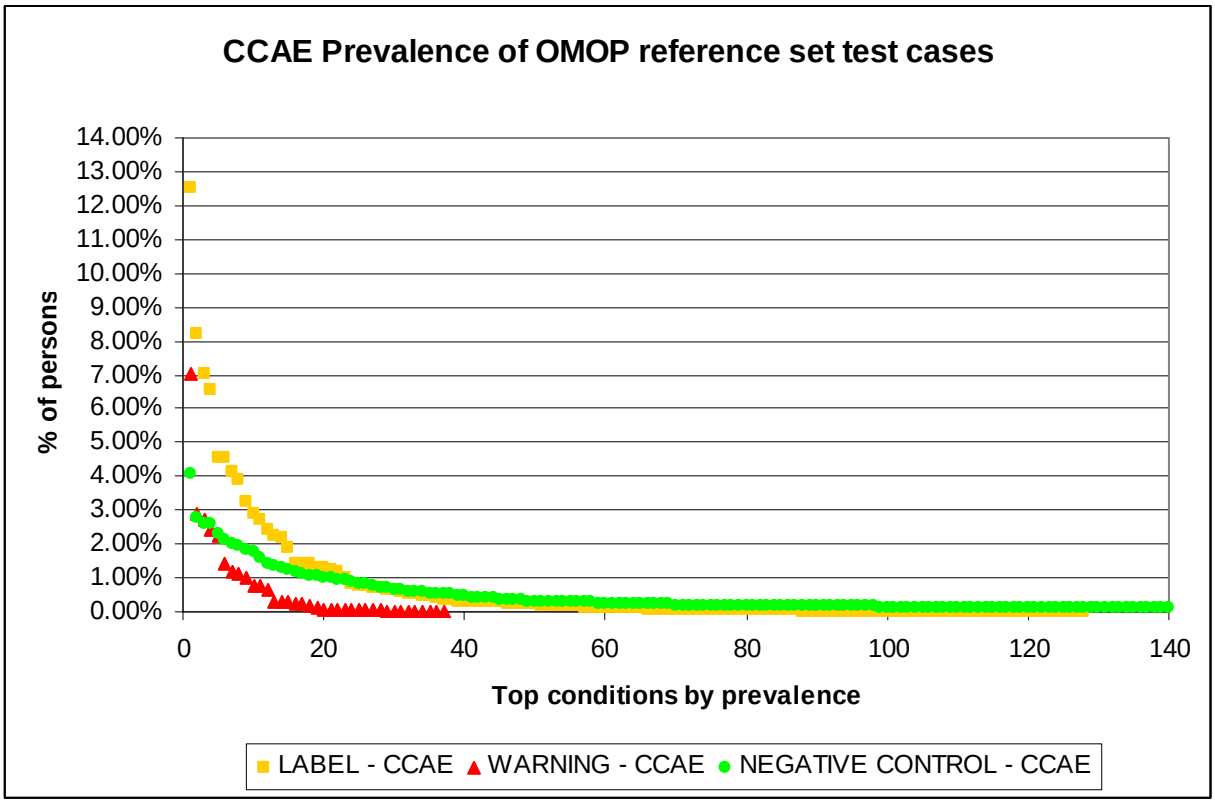

Figure 2: Prevalence of OMOP reference set test cases

It is recognized that the reference set of 'true associations' is not perfect. Because 'truth' is not known for any drug, we are required to select some surrogate (which has its own undefined sensitivity and specificity). We understand that labeled events have not necessarily been shown to be causally related to drug, or may not be expected to be observed in subsequent study. In particular, adverse events listed in the Adverse Reactions and Post-Marketing Experience section may reflect occurrence from clinical trials or spontaneous reporting without any expectation of causality. Similarly, it is possible that 'negative controls' have been selected that do have legitimate temporal relationships with the drugs of interest, and either have not been previously identified or were not listed on the product label.

The process used to identify the reference set was empirically driven to minimize subjective assessment, but carries its own limitations. SPLICER may misclassify adverse events, either missing or failing to code events that exist of the label or identifying terms on the labels that are not actual adverse events. While the application has strong performance characteristics for the Adverse Reactions section, it may be more prone to error in the Black Box or Warnings/Precautions sections due to the unstructured nature of the text. SPLICER classifies all matched terms meeting its criteria as potential adverse events, though may misclassify terms that were instead risk factors or contraindications.

This analysis of the structured product labels identified significant variability in labeled events, both within labels for the same ingredient and across ingredients. This analysis reinforces the

observation that the OMOP drugs of interest are not homogenous entities, even though they necessarily will be treated as such for purposes of experimentation. For example, OMOP is studying one concept of 'antiepileptics' as defined as exposure to either carbamazepine or phenytoin, but in practice, these two treatments have differing safety profiles. As a result, the heuristic of classifying labeled events as having occurred on >=50% of SPLs within the DOI was to account for this variability, but may introduce misclassification in the 'true associations' selected and may have unnecessarily restricted the test cases from other candidate conditions that may have provided additional exemplars.

That said, it is not necessary to identify all potential 'true associations' or all eligible 'negative controls'. Instead, the number of test cases can be considered the sample size within this methodological experiment. Because the same set of test cases is being consistently applied across all methods, any misclassification of test cases (either 'true associations' that are not related, or 'negative controls' that have an association) should not introduce differential bias to the experiment and should not influence the relative assessment of Mean Average Precision scores between methods.

Because the performance characteristics calculated are based on the artificial definition of truth used for experiment, care should be taken when attempting to predict how methods may perform prospectively in an active surveillance network. Instead, these metrics should be considered to be most appropriate for comparative purposes across methods and databases. Supplemental analyses using simulated data should provide additional insight into expected performance.

Acknowledgements:

Special thanks to Jon Duke and Jeff Friedlin of the Regenstrief Institute for development of SPLICER and their thoughtful insights into appropriate use of structured product label data informing OMOP's active surveillance research. Thanks also to Aaron Katz of University of North Carolina for supporting validation of SPLICER and for reviewing labels in support of developing the selection criteria proposed above.

References:

1. Miller DR, Oliveria SA, Berlowitz DR, et al. Angioedema Incidence in US Veterans Initiating Angiotensin-Converting Enzyme Inhibitors. Hypertension 2008.

2. Brown NJ, Snowden M, Griffin MR. Recurrent angiotensin-converting enzyme inhibitor--associated angioedema. JAMA 1997;278:232-3.

3. Brown NJ, Ray WA, Snowden M, et al. Black Americans have an increased rate of angiotensin converting enzyme inhibitor-associated angioedema. Clin Pharmacol Ther 1996;60:8-13.

4. Luthi JC, Lund MJ, Sampietro-Colom L, et al. Readmissions and the quality of care in patients hospitalized with heart failure. Int J Qual Health Care 2003;15:413-21.

5. Ahmed A, Maisiak R, Allman RM, et al. Heart failure mortality among older Medicare beneficiaries: association with left ventricular function evaluation and angiotensin-converting enzyme inhibitor use. South Med J 2003;96:124-9.

6. Johnson D, Jin Y, Quan H, et al. Beta-blockers and angiotensin-converting enzyme inhibitors/receptor blockers prescriptions after hospital discharge for heart failure are associated with decreased mortality in Alberta, Canada. J Am Coll Cardiol 2003;42:1438-45.

7. Luthi JC, McClellan WM, Fitzgerald D, et al. Mortality associated with the quality of care of patients hospitalized with congestive heart failure. Int J Qual Health Care 2002;14:15-24.

8. Schade CP, Hannah KL, Rezek G, et al. ACE inhibitor use and mortality reduction: do controlled trial results equal clinical practice? W V Med J 2006;102:304-6.

9. Sueta CA, Schenck A, Chowdhury M, et al. Effect of angiotensin-converting enzyme inhibitor therapy on 30-day outcome in patient > or = 65 years of age with chronic congestive heart failure. Am J Cardiol 2000;86:1151-3, A9.

10. Trewet CB, Shireman TI, Rigler SK, et al. Do ACE Inhibitors/Angiotensin II type 1 receptor antagonists reduce hospitalisations in older patients with heart failure? A propensity analysis. Drugs Aging 2007;24:945-55.

11. Bates DW, Su L, Yu DT, et al. Correlates of acute renal failure in patients receiving parenteral amphotericin B. Kidney Int 2001;60:1452-9.

12. Bates DW, Su L, Yu DT, et al. Mortality and costs of acute renal failure associated with amphotericin B therapy. Clin Infect Dis 2001;32:686-93.

13. Carson JL, Strom BL, Duff A, et al. Acute liver disease associated with erythromycins, sulfonamides, and tetracyclines. Ann Intern Med 1993;119:576-83.

14. Heaton PC, Fenwick SR, Brewer DE. Association between tetracycline or doxycycline and hepatotoxicity: a population based case-control study. J Clin Pharm Ther 2007;32:483-7.

15. de Abajo FJ, Montero D, Madurga M, et al. Acute and clinically relevant drug-induced liver injury: a population based case-control study. Br J Clin Pharmacol 2004;58:71-80.

18. Handoko KB, Souverein PC, van Staa TP, et al. Risk of aplastic anemia in patients using antiepileptic drugs. Epilepsia 2006;47:1232-6.

19. Wang PS, Bohn RL, Glynn RJ, et al. Zolpidem use and hip fractures in older people. J Am Geriatr Soc 2001;49:1685-90.

20. Wang PS, Bohn RL, Glynn RJ, et al. Hazardous benzodiazepine regimens in the elderly: effects of half-life, dosage, and duration on risk of hip fracture. Am J Psychiatry 2001;158:892-8.

21. Ray WA, Griffin MR, Downey W. Benzodiazepines of long and short elimination half-life and the risk of hip fracture. JAMA 1989;262:3303-7.

22. Ray WA, Griffin MR, Schaffner W, et al. Psychotropic drug use and the risk of hip fracture. N Engl J Med 1987;316:363-9.

23. Miller RR, Li YF, Sun H, et al. Underuse of cardioprotective medications in patients prior to acute myocardial infarction. Am J Cardiol 2003;92:209-11.

24. Krumholz HM, Radford MJ, Wang Y, et al. Early beta-blocker therapy for acute myocardial infarction in elderly patients. Ann Intern Med 1999;131:648-54.

25. Heidenreich PA, McClellan M. Trends in treatment and outcomes for acute myocardial infarction: 1975-1995. Am J Med 2001;110:165-74.

26. Fernandes AW, Madhavan SS, Amonkar MM, et al. Outcomes of inappropriate prescribing of beta-blockers after an acute myocardial infarction in a Medicaid population. Ann Pharmacother 2005;39:1416-22.

27. Chen J, Marciniak TA, Radford MJ, et al. Beta-blocker therapy for secondary prevention of myocardial infarction in elderly diabetic patients. Results from the National Cooperative Cardiovascular Project. J Am Coll Cardiol 1999;34:1388-94.

28. Chan AW, Quinn MJ, Bhatt DL, et al. Mortality benefit of beta-blockade after successful elective percutaneous coronary intervention. J Am Coll Cardiol 2002;40:669-75.

29. Barron HV, Viskin S, Lundstrom RJ, et al. Beta-blocker dosages and mortality after myocardial infarction: data from a large health maintenance organization. Arch Intern Med 1998;158:449-53.

35. Ettinger B, Pressman A, Schein J. Clinic visits and hospital admissions for care of acid-related upper gastrointestinal disorders in women using alendronate for osteoporosis. Am J Manag Care 1998;4:1377-82.

36. Donahue JG, Chan KA, Andrade SE, et al. Gastric and duodenal safety of daily alendronate. Arch Intern Med 2002;162:936-42.

37. Halkin H, Dushenat M, Silverman B, et al. Brand versus generic alendronate: gastrointestinal effects measured by resource utilization. Ann Pharmacother 2007;41:29-34.

38. Tata LJ, West J, Smith C, et al. General population based study of the impact of tricyclic and selective serotonin reuptake inhibitor antidepressants on the risk of acute myocardial infarction. Heart 2005;91:465-71.

39. Cohen HW, Gibson G, Alderman MH. Excess risk of myocardial infarction in patients treated with antidepressant medications: association with use of tricyclic agents. Am J Med 2000;108:2-8.

40. Ray WA, Meredith S, Thapa PB, et al. Cyclic antidepressants and the risk of sudden cardiac death. Clin Pharmacol Ther 2004;75:234-41.

41. Enger C, Weatherby L, Reynolds RF, et al. Serious cardiovascular events and mortality among patients with schizophrenia. J Nerv Ment Dis 2004;192:19-27.

42. Nakagawa S, Pedersen L, Olsen ML, et al. Antipsychotics and risk of first-time hospitalization for myocardial infarction: a population-based case-control study. J Intern Med 2006;260:451-8.

43. Liperoti R, Gambassi G, Lapane KL, et al. Conventional and atypical antipsychotics and the risk of hospitalization for ventricular arrhythmias or cardiac arrest. Arch Intern Med 2005;165:696-701.

44. Johnson SG, Rogers K, Delate T, et al. Outcomes associated with combined antiplatelet and anticoagulant therapy. Chest 2008;133:948-54.

45. Wysowski DK, Nourjah P, Swartz L. Bleeding complications with warfarin use: a prevalent adverse effect resulting in regulatory action. Arch Intern Med 2007;167:1414-9.

46. Zhang K, Young C, Berger J. Administrative claims analysis of the relationship between warfarin use and risk of hemorrhage including drug-drug and drug-disease interactions. J Manag Care Pharm 2006;12:640-8.

47. Boulanger L, Hauch O, Friedman M, et al. Warfarin exposure and the risk of thromboembolic and major bleeding events among medicaid patients with atrial fibrillation. Ann Pharmacother 2006;40:1024-9.

48. Hollowell J, Ruigomez A, Johansson S, et al. The incidence of bleeding complications associated with warfarin treatment in general practice in the United Kingdom. Br J Gen Pract 2003;53:312-4.

