Drug Saf (2013) 36 (Suppl 1):S15–S25
DOI 10.1007/s40264-013-0103-1

ORIGINAL RESEARCH ARTICLE

Variation in Choice of Study Design: Findings from the Epidemiology Design Decision Inventory and Evaluation (EDDIE) Survey

Paul E. Stang · Patrick B. Ryan · J. Marc Overhage · Martijn J. Schuemie · Abraham G. Hartzema · Emily Welebob

Abstract

Background Researchers using observational data to understand drug effects must make a number of analytic design choices that suit the characteristics of the data and the subject of the study. Review of the published literature suggests that there is a lack of consistency even when addressing the same research question in the same database.

Objective To characterize the degree of similarity or difference in the method and analysis choices made by observational database research experts when presented with research study scenarios.

Research Design On-line survey using research scenarios on drug-effect studies to capture method selection and analysis choices that follow a dependency branching based on response to key questions.

Subjects Voluntary participants experienced in epidemiological study design solicited for participation through registration on the Observational Medical Outcomes Partnership website, membership in particular professional organizations, or links in relevant newsletters.

Measures Description (proportion) of respondents selecting particular methods and making specific analysis choices based on individual drug-outcome scenario pairs. The number of questions/decisions differed based on stem questions of study design, time-at-risk, outcome definition, and comparator.

Results There is little consistency across scenarios, by drug or by outcome of interest, in the decisions made for design and analyses in scenarios using large healthcare databases. The most consistent choice was the cohort study design but variability in the other critical decisions was common.

The OMOP research used data from Truven Health Analytics (formerly the Health Business of Thomson Reuters), and includes MarketScan® Research Databases, represented with MarketScan Lab Supplemental (MSLR, 1.2 m persons), MarketScan Medicare Supplemental Beneficiaries (MDCR, 4.6 m persons), MarketScan Multi-State Medicaid (MDCD, 10.8 m persons), MarketScan Commercial Claims and Encounters (CCAE, 46.5 m persons). Data also provided by Quintiles® Practice Research Database (formerly General Electric's Electronic Health Record, 11.2 m persons) database. GE is an electronic health record database while the other four databases contain administrative claims data.

Electronic supplementary material The online version of this article (doi:10.1007/s40264-013-0103-1) contains supplementary material, which is available to authorized users.

P. E. Stang · P. B. Ryan
Janssen Research and Development LLC,
Titusville, NJ, USA

P. E. Stang (✉)
1125 Trenton-Harbourton Road,
Room K30205, 200,
Titusville, NJ 08560, USA
e-mail: pstang@its.jnj.com

J. M. Overhage
Siemens Health Services, Malvern, PA, USA

M. J. Schuemie
Department of Medical Informatics, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

A. G. Hartzema
College of Pharmacy, University of Florida, Gainesville, FL, USA

P. E. Stang · P. B. Ryan · J. M. Overhage · M. J. Schuemie · E. Welebob
Observational Medical Outcomes Partnership, Foundation for the National Institutes of Health, Bethesda, MD, USA

Conclusions There is great variation among epidemiologists in the design and analytical choices that they make when implementing analyses in observational healthcare databases. These findings confirm that it will be important to generate empiric evidence to inform these decisions and to promote a better understanding of the impact of standardization on research implementation.

1 Background

The advent of large claims and electronic health record databases has given rise to a number of applications in epidemiology, health economics, comparative effectiveness, and health services research. However, the availability of these data has also introduced a number of complexities into the key analytic decisions for the study of the safety of medical interventions. As these data were collected for reasons other than research, these data often lack the features, detail, or data accuracy that would be desirable and treatment has not been assigned randomly increasing the likelihood of confounding [1, 2]. There are many ways to attempt to mitigate these limitations, leaving the researcher to make analytic design choices that suit the characteristics of the data and the subject of the study. For example, in certain situations, using a 'self-controlled design' such as the self-controlled case series (SCCS) [3] might be best to limit the effect of between-subject confounding. On the other hand, case-control studies, developed in an era where the cost of data collection or the time required was a critical concern, are often used in the study of rare outcomes or as an exploratory approach where little is known about the association between the exposure and health outcome of interest [4]. While case-control studies provide only relative estimates of effect (relative risk, rates), cohort studies allow for a direct calculation of these effects as well as providing absolute measures of incidence. Cohort designs are also intuitively easier to understand and explain as they track with the passage of time and more closely resemble other research study designs, particularly the clinical trial. Cohort designs are also more favorably rated in the hierarchy of evidence used in assessing clinical evidence [5]. Decisions are required not only about the general method (e.g. cohort vs. case-control vs. self-controlled case series) but also on the definition of the Health Outcomes of Interest (HOI), and depending on the method, choice of comparator, the duration of follow-up during which events are "counted" as associated with a particular exposure, and the selection of covariates for adjustment among others [1, 2].

In previous work [6] testing 14 analytic methods, the number of different combinations of analysis choices (eg, 'parameter' decisions) varied from 48 (for case-control and case-crossover designs) to 162 (for observational screening) and included definition of time-at-risk, identification of HOIs based on first occurrence or all occurrences of diagnosis codes, choice of comparator group, exposure and HOI definitions, and specific confounding adjustment strategy. This spectrum of choice, even within the same method, suggests that there is ample opportunity across researchers to differ in their decisions of how to implement a method in a healthcare database. Our study of the variation in choice of HOI definitions [7] confirms this but systematically reviewing publications of database studies, even for the same research question, will not capture all of the possible decisions nor all of the analyses that were actually performed.

In order to characterize the degree of similarity or difference in the method and analysis choices, we sought to determine what choices observational database research experts would make in a series of reasonably clearly articulated scenarios. A greater understanding of the choices that are made among a group of experienced scientists can help inform an active drug surveillance system as well as any validation and standardization work that will be required.

2 Methods

2.1 Study Design

An online survey format was used to reach a broad participant pool and to maximize the likelihood of participation. This study was approved by the Indiana University Institutional Review Board.

2.2 Survey Development

The questions were developed by the Observational Medical Outcomes Partnership (OMOP) research team and the survey was fielded using SurveyGizmo™. This tool allows for the use of branching questions based on respondent answers, as well as allowing the respondent to pause during a survey, save their answers, and return at a later time to complete the survey which, given the content and length of the survey, was particularly important. The survey was piloted with 10 individuals for quality and clarity checks, as well as to identify any logistical problems.

2.3 Survey Content

The initial screening required the respondent be able to read and write English and actively consent to participation. There were no quotas sought for the sample and no *a priori* sample size estimates were determined. An initial

series of questions captured respondent's academic background, relevant years of experience, and experience with the particular databases, medications, and HOIs which were used to categorize expertise. Respondents were randomized into one of the five groups of 6 drug-HOI pairs, each covering 2 different drugs and 3 different HOIs, representing two positive controls (ie, drug-HOI pairs believed to describe a true adverse drug reaction) and four negative controls (ie. drug-HOI pairs where no causal relationship is believed to exist) (see http://omop.org). This approach of using known positive and negative associations is part of the overall OMOP program of research and in the present case, allowed us to examine response patterns for potential associations that would not be familiar to participants. Respondents were presented with a specific drug-HOI pair and asked a series of questions about how they would design an epidemiology evaluation study. The pairs were selected from the positive and negative controls already identified in the OMOP program of research [8]. After completing the first group of drug-HOI pairs, respondents were invited to repeat the exercise for as many of the 4 other groups of 6 drug-HOI pairs as they wished. The specific questions asked for each drug-HOI pair are provided in the "Appendix". The five groups were:

Group 1: Drugs: Angiotensin-converting-enzyme [9] inhibitors, Antiepileptics HOIs: Angioedema, Aplastic anemia, Gastrointestinal (GI) ulcer hospitalization

Group 2: Drugs: Benzodiazepine, Warfarin HOIs: Hip fracture, Bleeding, Aplastic anemia

Group 3: Drugs: Antibiotics, Bisphosphonates HOIs: Acute liver injury, Gastrointestinal (GI) ulcer hospitalization, Acute myocardial infarction (MI)

Group 4: Drugs: Tricyclic antidepressants, Amphotericin B HOIs: Acute myocardial infarction, Acute renal failure, Acute liver injury

Group 5: Drugs: Typical antipsychotics, Bisphosphonates HOIs: Acute myocardial infarction, Acute renal failure, GI ulcer hospitalization

The questions for each drug-HOI pair varied by responses and question branching based on design features of a given method chosen but were consistent across all drug-HOI pairs. The choices and branching were designed to reflect actual practice as exemplified in the many different ways to define an HOI. For every drug-HOI pair, we asked the participants to select one of several possible HOI definitions. For acute myocardial infarction, for example, the respondent was given the option of selecting the occurrence of a single diagnostic code (ICD-9 CM 410* 'Acute myocardial infarction'), occurrence of one of multiple codes (ICD-9 CM 410* 'Acute myocardial infarction', 411.1 'Intermediate coronary syndrome', 411.8 'Other acute coronary occlusion'), or the occurrence of a combination of diagnosis and procedure codes that included time windows from the diagnosis code. The full list of HOI definitions used can be found at http://omop.org/HOI [7]. Completion of one drug-HOI pair was followed by initiation of the next drug-HOI pair in the sequence. The questionnaire for each drug-HOI pair was expected to take approximately 10 min.

2.4 Subject Recruitment

We employed a wide variety of methods to recruit participants. Solicitation emails were sent through OMOP to all registrants of the OMOP public website (http://omop.org) (approximately 1,500). Recipients were encouraged to forward the invitation to their staff, colleagues and other contacts that may be interested but not registered on the OMOP website. Similar solicitation e-mails or newsletter links were distributed through the Brookings Institute, the International Society for Pharmacoepidemiology (ISPE), and by individual researchers active in OMOP. A link to the survey was also posted on the OMOP website, Foundation for the National Institutes of Health (FNIH) website, and LinkedIn. Once on the survey website, the respondent was required to endorse a statement of agreement to participate in the survey before initiating the survey.

Because of the time and potential complexity of this survey, two levels of incentives were available by lottery. Respondents who complete all 6 pairs within a group were entered into a drawing for Prize #1: a $100.00 Amazon gift card. This prize was drawn for every 50 respondents who complete the initial 6 pairs so the chance of winning was 1/50; completion of all 5 groups (30 drug-HOI pairs) entered the respondent into a drawing for Prize #2: a $1,000.00 Amazon gift card which was drawn for every 25 respondents who completed all 5 groups so the chance of winning at this level is 1/25.

2.5 Data analysis

The survey system captured the data directly as the respondent answered the survey questions and was later exported into SAS datasets for analysis. An e-mail address was the only personally identifiable information collected to allow saving and re-entry into the survey, reminders to encourage survey completion, and to notify the prize winners. E-mail addresses were not available to the data analyst. All analyses conducted on the survey results were descriptive, reported as population-level summaries across all respondents or specific subpopulations of the respondents based on their self-reported educational/professional background or stated experience with specific types of study methods, drugs, or HOIs. In order to stratify results for analysis, we defined 'experts' as respondents meeting the following 3 self-reported criteria: a doctoral-level

degree (PhD, ScD or MD), either a graduate degree in Epidemiology or current job position in Epidemiology, and having either been a principal investigator or lead the design of an observational database study.

For each drug-HOI pair, multiple choice questions were summarized as categorical variables by the proportion of respondents within each response category. Question responses were compared across drug-HOI pairs within respondent to determine which questions yield different answers when the drug and/or HOI are changed and to identify which, if any, questions generate consistent responses that are independent of the drug and/or HOI. We focused our analyses on the 4 HOIs that are of particular interest in drug safety and formed the basis of subsequent research in OMOP: acute kidney injury, acute liver injury, acute myocardial infarction, and gastrointestinal ulcer hospitalization.

3 Results

The survey was fielded October 15–December 31, 2011. At the close of the survey, 136 individuals had completed at least one drug-HOI pair of whom 79 met our criteria for 'expert' (Table 1). Because of the open solicitation through a number of channels that overlap and the unknown 'expert' status of non-responders, a response rate could not be estimated. In total, respondents completed 721 drug-HOI scenarios that were used to assess consistency and agreement.

Respondents classified as 'experts' averaged 5.7 studies conducted in the past 12 months, with 76 % of 'expert' respondents reporting work on at least 3 studies. 'Experts' had 11 years since graduation on average, with 66 % of those respondents having more than 5 years since their last degree. For all drug-outcome pairs, fewer than 25 % of respondents reported having prior experience conducting a study of the target drug and the target outcome of interest.

Table 1 Respondent background (Categories are not mutually exclusive)

Respondents	N
Total	**136**
Has MD or PhD	101
Has Epidemiology degree	87
Has Epidemiology position	86
Has principal investigator experience	80
Has study design experience	105
'Expert' : 1) has PhD or MD or DSc, 2) either had graduate degree in Epidemiology or current position in Epidemiology, and 3) has either been principal investigator or lead design on observational database study	79

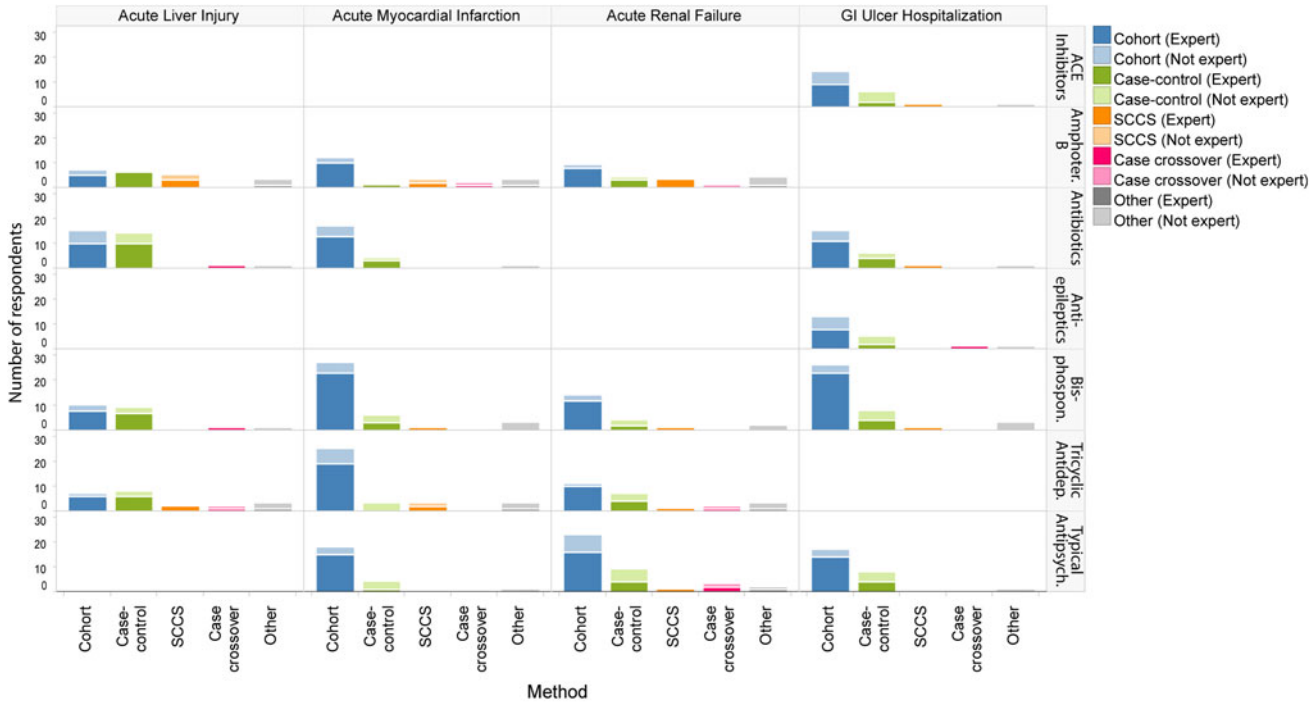
As a result, complete subgroup analyses among respondent reporting past domain experience were not completed. However, within the small samples available, heterogeneity in response was observed. For example, among the five 'expert' respondents who reported experience studying antibiotics and acute liver injury who were asked to design a study of the potential antibiotic-liver injury effect, three selected a cohort design, one selected a case-control design, and one selected a case-crossover design. Across all drug-outcome combinations with 2 or more respondents selecting the new user cohort design, there was lack of concordance in the primary confounding adjustment strategy to be used. At least one respondent selected each of available confounding adjustment options, which included stratification by age and sex, multivariate logistic regression, propensity score stratification using 5 strata, propensity score stratification using 20 strata, including the propensity score as covariate in the outcome model. Additionally, respondents specified other strategies, including propensity score matching and marginal structural modeling.

The study method choices made by respondents are reflected in Fig. 1. For most drug-HOI pairs, the cohort method was favored over the other methods, most pronounced in the study of acute myocardial infarction and GI ulcer hospitalization regardless of drug. Acute liver injury by contrast showed a more even distribution of study method choice as did many of the combinations for acute renal failure. The distribution of choice of method by experts followed a similar pattern, although experts were seemingly more willing to implement self-controlled and case-crossover methods than non-experts. It is noteworthy that, despite a modest endorsement of a particular method by a substantial proportion of respondents, the results do not reflect a consensus nor the appearance of near consensus in any of the drug-HOI pairs presented.

Intra-respondent consistency in study method selection within a given HOI was also assessed (Table 2). Most striking is that 71 % of respondents would choose the same study method (cohort) to study GI ulcer, regardless of drug involved. Acute kidney injury was the HOI least likely to be studied across drug exposure with the same study method.

Choice of preferred HOI definition also shows little consistency among respondents regardless of the HOI being studied or the number of possible definitions (Fig. 2). GI Ulcer hospitalization definition choice (only 3 HOI definition choices) was evenly divided between two of the definitions while acute liver injury, acute MI, and acute renal failure varied across medication and across definitions. Variability increased with increasing number of HOI definition choices.

The critical decision of the 'Time at Risk' window choice is depicted in Fig. 3. While 30 days after the end of



HOI = health outcome of interest; **SCCS** = self-controlled case series

Fig. 1 Frequency of Analytic Method Choice by Drug-HOI Pair by 'expert' (*darker colors*) and 'non-expert' (*lighter colors*)

Table 2 Intra-respondent Consistency of Method Choice

Outcome	Total Number of Respondents	% of respondents who selected same method for all drugs
Acute Kidney Injury	29	45 %
Acute Liver Injury	31	65 %
Acute Myocardial Infarction	34	62 %
Gastrointestinal Ulcer	38	71 %

the exposure era was the most frequent choice, particularly within cohort method, there is some variation by both drug and HOI, notably the prominence of 'all post-exposure time' for acute MI regardless of drug and tricyclic antidepressants across 3 of the 4 HOIs.

Among 'expert' respondents who selected new user cohort designs, there was substantial variation in choice of confounding adjustment strategy. Among the 22 respondents designing a cohort design for bisphosphonates and GI ulcer, less than one third of respondents agreed to the same confounding adjustment strategy: 7 chose propensity score stratification with 5 strata, 4 chose propensity score matching, 3 chose propensity score stratification using 20 strata, 3 chose multivariate logistic regression, and 3 chose to include the propensity score as a covariate in the outcome model.

As Table 3 shows, there was little agreement amongst 'expert' respondents regarding which of 10 types of databases they would most likely use in analysis of each of the drug-outcome pairs. Among the 19 respondents evaluating ACE inhibitor and angioedema, 58 % thought aggregated employer-based administrative claims would be used while 42 % did not select this data as applicable. In only 39 cases among the 280 drug-outcome-database combinations evaluated (14 %) did >80 % of respondents agree that a data source was mostly likely to be used for that analysis.

Health information exchanges like Regenstrief Institute and clinical records from Veterans Affairs were the only two types of data which were selected by a majority of respondents for all 28 drug-outcome pairs evaluated. Administrative claims data captured at the point of care (e.g., Surveillance Data Inc), was the least likely type of data to be selected, on average, but still 36 % of respondents deemed it appropriate for the pairs of interest. The respondents were mostly divided on the potential utility of data from a consortium of providers using a common electronic health record (e.g., GE Centricity), with 19 of 28 drug-outcome pairs having a proportion of that data type selected between 40 % and 60 %.

Figure 4 provides a clear example of a single drug-HOI pair (bisphosphonates and GI ulcer hospitalization) and the distribution of survey responses over 4 levels of

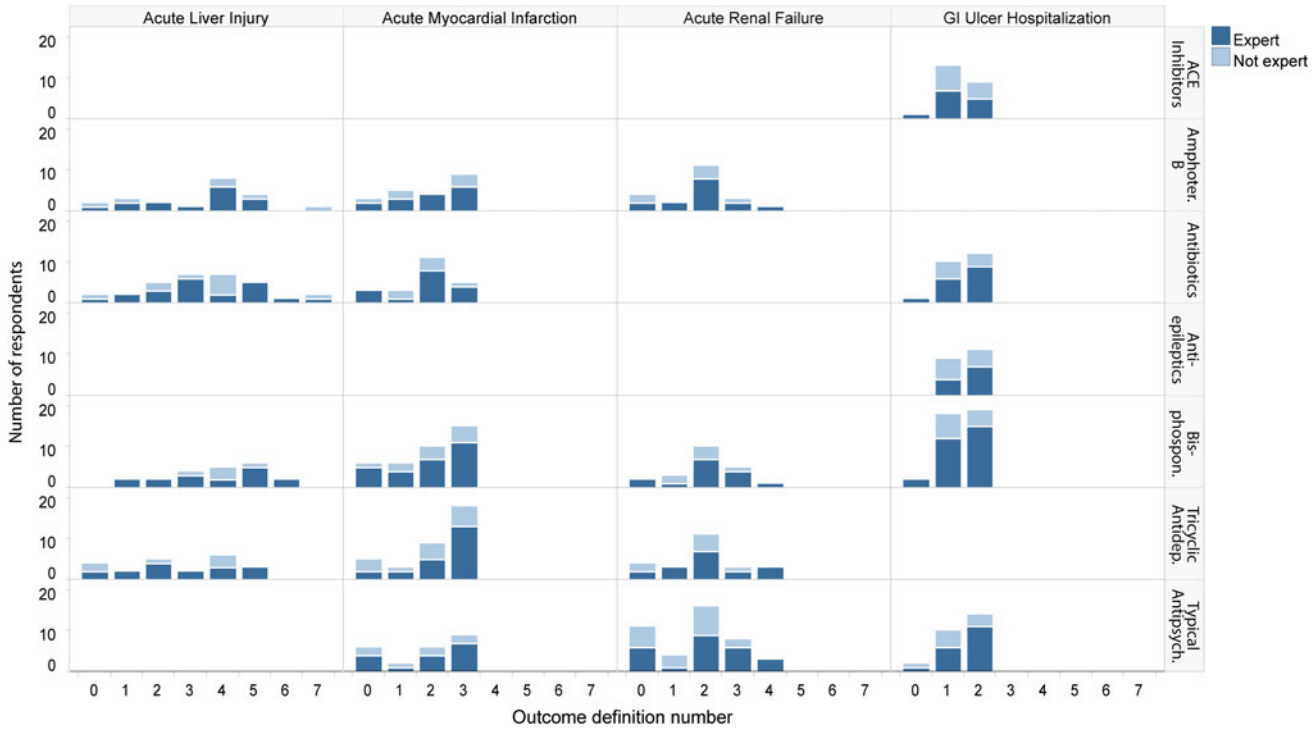
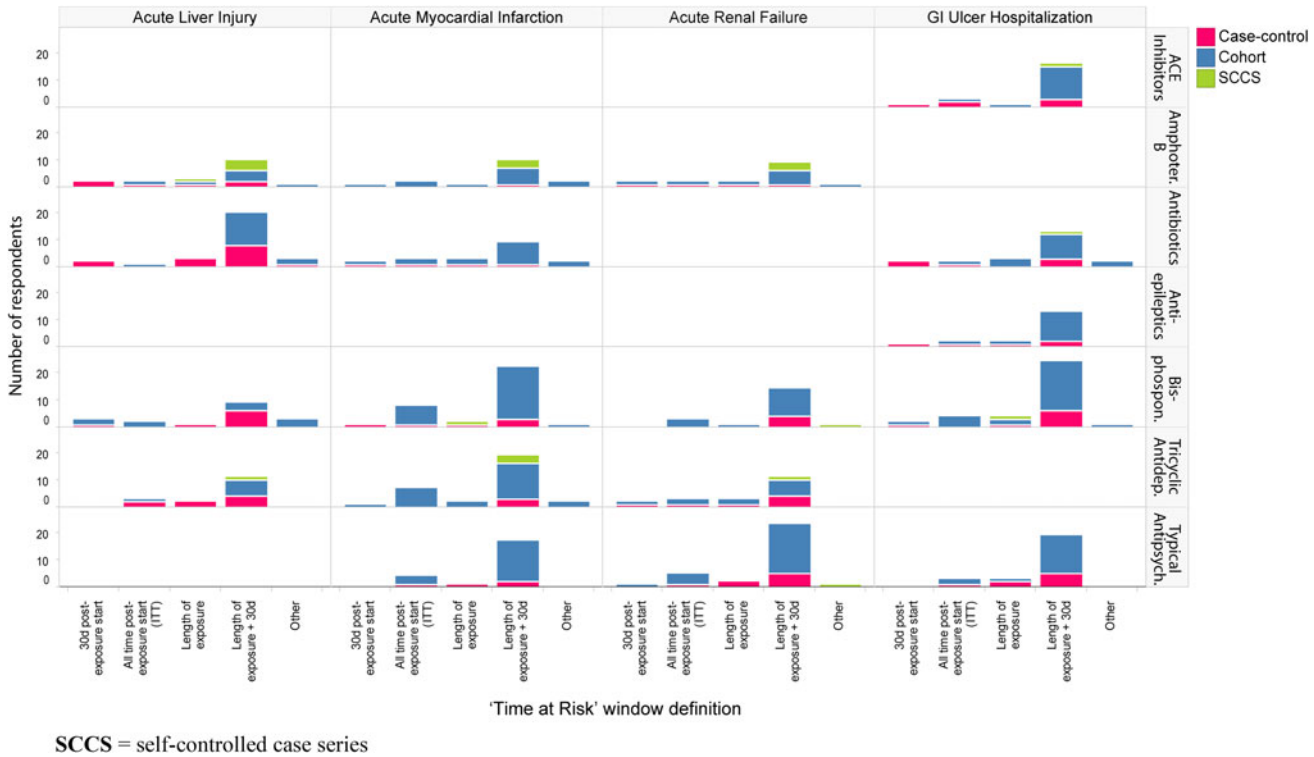


Fig. 2 Choice of health outcome of interest definition by 'expert' and non-expert



SCCS = self-controlled case series

Fig. 3 'Time at Risk' window choice by drug-health outcome of interest pair

questions. Most striking is the divergent selections with each dependent decision parsing into at least 2 selections for each subsequent decision. A selection of one of 3 HOI definitions quickly gives rise to 15 comparator-method combinations after 3 steps in the decision-making process.

Table 3 Database selection by experts across drug-HOI combinations

Drug	Outcome	# of 'expert' respondents	Aggregated employer-based administrative claims, privately insured population	Administrative claims from commercially-insured population with supplemental Medicare insurance	Multi-state Medicaid administrative claims database	Administrative claims population supplemented with laboratory values	Consortium of general practitioners and specialists using same electronic health record system but not necessarily part of the same delivery system	HIE (electronic capture of healthcare information across organizations within a region or community)	Regional healthcare system / Integrated delivery network (eg, regional HMO)	Administrative claims from large payer	Administrative claims captured at the point of care (not unique to a particular insurance plan) (eg, SDI Health)	Clinical and administrative records for Veteran population covered by a federal agency
ACE Inhibitor	Angioedema	19	58 %	58 %	53 %	42 %	47 %	53 %	53 %	58 %	26 %	74 %
ACE Inhibitor	Aplastic Anemia	14	36 %	43 %	43 %	79 %	36 %	71 %	64 %	57 %	21 %	71 %
ACE Inhibitor	GI Ulcer Hospitalization	13	69 %	54 %	38 %	46 %	62 %	69 %	69 %	62 %	38 %	85 %
Amphotericin B	Acute Liver Injury	15	67 %	80 %	60 %	80 %	40 %	80 %	73 %	67 %	40 %	87 %
Amphotericin B	AMI	15	73 %	67 %	60 %	60 %	47 %	87 %	73 %	73 %	53 %	87 %
Amphotericin B	Acute Renal Failure	15	60 %	67 %	60 %	73 %	47 %	87 %	80 %	60 %	47 %	80 %
Antibiotics	Acute Liver Injury	21	24 %	24 %	33 %	71 %	52 %	57 %	81 %	62 %	48 %	86 %
Antibiotics	AMI	16	75 %	69 %	63 %	75 %	44 %	63 %	63 %	63 %	31 %	81 %
Antibiotics	GI Ulcer Hospitalization	16	69 %	75 %	69 %	81 %	44 %	75 %	75 %	75 %	44 %	94 %
Antiepileptics	Angioedema	12	67 %	50 %	50 %	50 %	67 %	67 %	75 %	75 %	50 %	83 %
Antiepileptics	Aplastic Anemia	11	45 %	45 %	36 %	91 %	45 %	73 %	64 %	64 %	27 %	73 %
Antiepileptics	GI Ulcer Hospitalization	11	73 %	45 %	45 %	55 %	73 %	73 %	73 %	73 %	45 %	91 %
Benzodiazepines	Aplastic Anemia	15	60 %	60 %	60 %	87 %	60 %	60 %	47 %	60 %	33 %	80 %
Benzodiazepines	Bleeding	22	64 %	73 %	64 %	68 %	68 %	64 %	77 %	59 %	41 %	77 %
Benzodiazepines	Hip fracture	25	64 %	84 %	64 %	56 %	56 %	64 %	56 %	68 %	40 %	68 %
Bisphosphonates	Acute Liver Injury	16	44 %	44 %	44 %	63 %	63 %	69 %	75 %	63 %	25 %	88 %
Bisphosphonates	AMI	27	78 %	81 %	67 %	74 %	59 %	81 %	85 %	78 %	59 %	96 %
Bisphosphonates	Acute Renal Failure	15	53 %	80 %	53 %	87 %	53 %	67 %	73 %	67 %	33 %	80 %
Bisphosphonates	GI Ulcer Hospitalization	29	69 %	79 %	59 %	59 %	52 %	69 %	79 %	76 %	48 %	90 %
Tricyclic Antidepressants	Acute Liver Injury	16	63 %	81 %	69 %	81 %	50 %	69 %	88 %	56 %	19 %	94 %
Tricyclic Antidepressants	AMI	22	64 %	68 %	45 %	59 %	45 %	73 %	77 %	55 %	50 %	73 %
Tricyclic Antidepressants	Acute Renal Failure	17	53 %	71 %	53 %	76 %	35 %	71 %	88 %	53 %	24 %	82 %
Typical Antipsychotics	AMI	16	69 %	81 %	81 %	69 %	63 %	75 %	81 %	75 %	44 %	88 %
Typical Antipsychotics	Acute Renal Failure	25	32 %	52 %	44 %	84 %	56 %	56 %	56 %	56 %	16 %	84 %

Table 3 continued

Drug	Outcome	# of 'expert' respondents	Aggregated employer-based administrative claims, privately insured population	Administrative claims from commercially-insured population with supplemental Medicare insurance	Multi-state Medicaid administrative claims database	Administrative claims population supplemented with laboratory values	Consortium of general practitioners and specialists using same electronic health record system but not necessarily part of the same delivery system	HIE (electronic capture of healthcare information across organizations within a region or community)	Regional healthcare system / Integrated delivery network (eg, regional HMO)	Administrative claims from large payer	Administrative claims captured at the point of care (not unique to a particular insurance plan) (eg, SDI Health)	Clinical and administrative records for Veteran population covered by a federal agency
Typical Antipsychotics	GI Ulcer Hospitalization	18	67 %	78 %	72 %	44 %	50 %	72 %	83 %	78 %	50 %	78 %
Warfarin	Aplastic Anemia	10	50 %	50 %	40 %	80 %	40 %	70 %	70 %	60 %	10 %	90 %
Warfarin	Bleeding	12	75 %	67 %	58 %	75 %	50 %	67 %	67 %	67 %	25 %	83 %
Warfarin	Hip fracture	12	83 %	67 %	58 %	67 %	42 %	67 %	67 %	67 %	25 %	83 %

AMI Acute Myocardial Infarction, *HIE* health information exchange

4 Discussion

We can draw two main conclusions from these results. First, there is not a plurality of epidemiologists in our sample who would make the same choices when implementing an analysis for the same drug-HOI pair. Second, recent empirical evaluations of the performance of study methods and method choices casts doubts on the appropriateness of the most common choices made by the respondents [10–20]. These findings are consistent with the literature and systematic reviews with observational data, where we also see inconsistency of analytic method choices across the study of the same HOI [7]. For example, a meta-analysis of observational studies of nonsteroidal anti-inflammatory drugs (NSAIDs) and myocardial infarction by Hernandez-Diaz et al. [21] included 16 studies, 15 of which were in healthcare databases. Four of these studies were cohorts while the remaining 12 were case-control studies (9 nested case-control studies). Loke and colleagues [22], in their meta-analysis of 16 studies of thiazolidinediones and MI included 4 case-control studies (all nested) and 12 cohort studies all in healthcare databases. Similarly, the literature (including lay press) is rife with conflicting studies of the same drug-HOI pair (sometimes using the same data source) whose results are contradictory yet differing by only one or two decisions in how a study was implemented. One only needs to examine the impact of subtle choices in some published studies. For example, the risk of esophageal cancer among users of oral bisphosphonates was studied by two different groups using the same database (the United Kingdom General Practice Research Database-General Practice Research Database) [23, 24]. Their studies differed in choice of study method (cohort vs. case-control), time-at-risk, matching criteria, and exclusions which resulted in one study determining as high as a 2-fold relatively higher risk vs. the second study finding no increased risk.

Our survey appears to be the first of its kind in attempting to document the choices made by an informed group of practitioners who have expertise in the implementation of epidemiologic studies in large healthcare databases. Although similar findings might be inferred from the published literature, meta-analyses, and personal experience, the current study sought to systematically investigate the extent to which there is agreement on the basic implementation choices that would be inclusive of a larger group of scientists and would not be restricted to those analyses appearing in the peer-reviewed literature.

Our results document the degree of consistency (or lack thereof) across choice of study method and HOI definition. One would have expected consistency across respondents when addressing the same drug-HOI combination and more

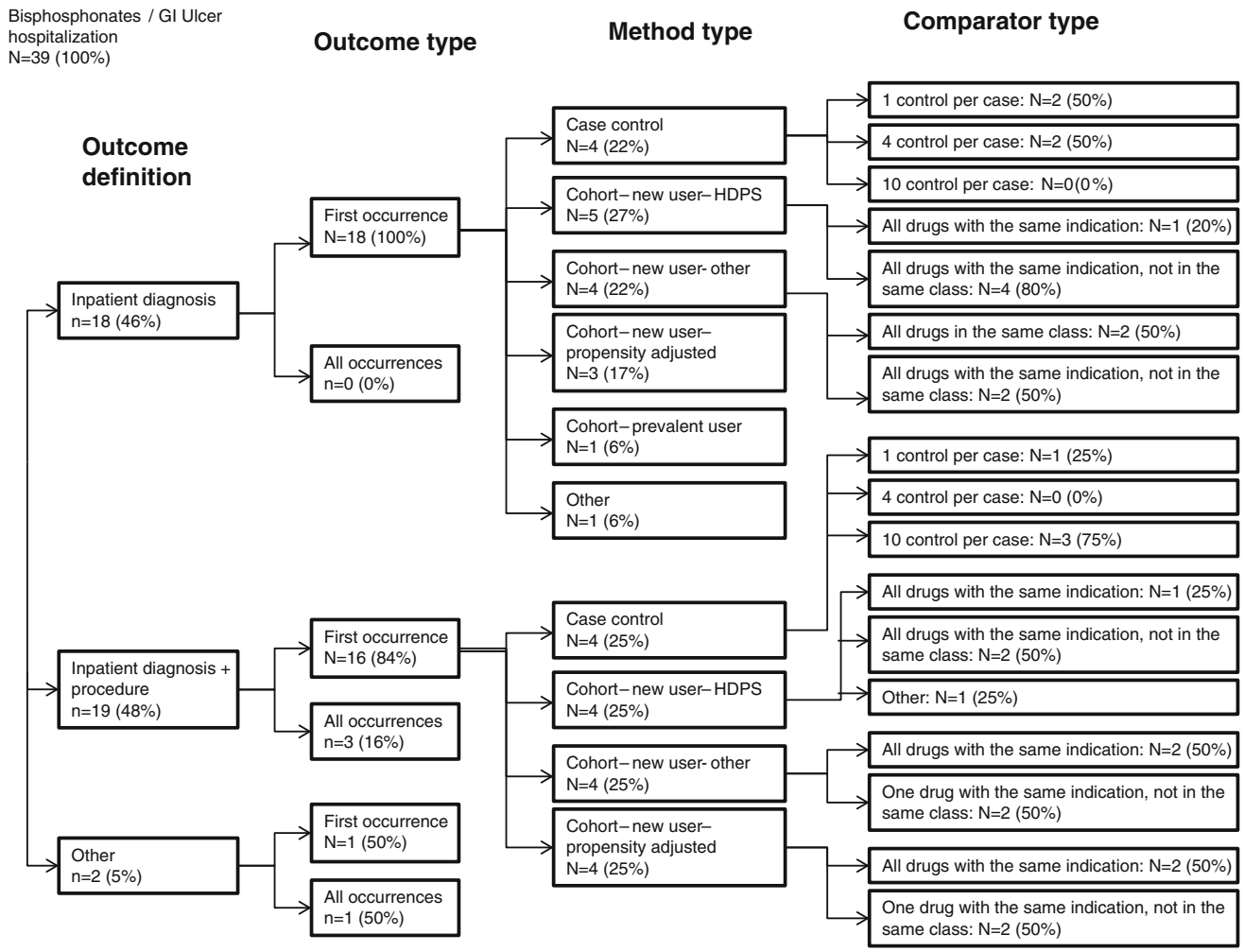


Fig. 4 Depiction of variation in 4 key choices in assessing Bisphosphonates and GI Ulcer Hospitalization (dependencies as reflected in the survey)

variation across these examples. Our findings show that the intra-respondent consistency for the same outcome across different drugs was very high but did differ somewhat by outcome. It is difficult to determine if this is because of comfort with a particular method, expediency in participating in the survey but it does suggest that the findings from research that may suggest better alternatives may be met with some resistance among experts. Although no real standard appears to exist, there have been recent efforts in pharmacoepidemiology to develop a structured approach to determining a study method based on the features of the exposure, HOI and source of potential confounding [25, 26]. However, our survey shows that we are far from implementing consistent approaches to design feature selection. Many of the decisions are based largely on non-empiric assessments of their impact, but efforts to measure the impact of some of these decisions have been undertaken [10]. Together, these conclusions suggest that the field needs to adopt a more systematic approach to choosing study designs, defining parameters, and implementing controls for potential confounding.

We attempted to capture insights on a number of key decisions made however we did not address all possibilities that may exist outside of method, database, and definitions. Our attempt to study control for potential confounding outside of the confounding adjustment embedded in many of the method and analysis choices is a case in point. Previous research by Vessey and others [27–29] has illustrated that people experiencing an event often have a predisposing or alternate proximal cause. In the context of the current research, those with a predisposing cause might be less likely than others to be using the drug of interest. The predisposing cause may be difficult to determine and attribute using existing data; controlling for these effects and/or considering them (or excluding subjects with them from the case definition) were not explicitly solicited in this survey but are considerations in the control of potential biases when assessing drug-outcome associations. The

work by Hernán et al. [30] has also shown that it is not just confounding but selection criteria regarding cases that can be very important, and selection bias of this type may be more important than confounding itself. Perrio et al. [31] have attempted to catalogue the use of 'exclusion criteria' in pharmacoepidemiologic studies and defined 5 broad categories related to data, disease, exposure, patient, and a miscellaneous category wherein they specifically detail some of the controversy surrounding the implementation of 'disease-related' exclusions as a way to control for some confounding.

Our study has other potential limitations. We did not have 100 % response from those invited to participate but did focus our analysis on those respondents deemed to be qualified, capable and therefore more likely to successfully undertake this type of research and communicate its findings to a decision-maker. While the sample size is modest, the findings of variability are still valuable. We also accept that the potentially artificial nature of asking someone to define a study design and general study without opportunity to fully review literature, prior studies, pharmacology, database characteristics, and possible confounders may not completely reflect how a study would be implemented.

5 Conclusions

There is great variability among epidemiologists in the design choices that they would make when implementing analyses in observational healthcare databases. Because of the growing importance of these data in determining the effects of medical interventions and in guiding regulatory actions, it will be important to inform these decisions to help standardize approaches with more empiric evidence. This standardization, and an understanding of the underlying impact of study decisions, could help make the evidence emanating from research in large healthcare databases more credible.

Acknowledgments The Observational Medical Outcomes Partnership is funded by the Foundation for the National Institutes of Health (FNIH) through generous contributions from the following: Abbott, Amgen Inc., AstraZeneca, Bayer Healthcare Pharmaceuticals, Inc., Bioden Idec, Bristol-Myers Squibb, Eli Lilly & Company, GlaxoSmithKline, Janssen Research and Development, Lundbeck, Inc., Merck & Co., Inc., Novartis Pharmaceuticals Corporation, Pfizer Inc, Pharmaceutical Research Manufacturers of America (PhRMA), Roche, Sanofi-aventis, Schering-Plough Corporation, and Takeda. Drs. Stang, Schuemie and Ryan are employees of Janssen Research and Development. Dr. Schuemie received a fellowship from the Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration and has become an employee of Janssen Research & Development since completing this research.

Drs. Schuemie and Madigan have received a grant previously from FNIH. J. Marc Overhage, Abraham G. Hartzema, and Emily Welebob have no conflicts of interest to declare.

This article was published in a supplement sponsored by the Foundation for the National Institutes of Health (FNIH). The supplement was guest edited by Stephen J.W. Evans. It was peer reviewed by Olaf H. Klungel who received a small honorarium to cover out-of-pocket expenses. S.J.W.E has received travel funding from the FNIH to travel to the OMOP symposium and received a fee from FNIH for the review of a protocol for OMOP. O.H.K has received funding for the IMI–PROTECT project.from the Innovative Medicines Initiative Joint Undertaking (http://www.imi.europa.eu) under Grant Agreement no 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in kind contribution.

References

1. Strom B, editor. Pharmacoepidemiology. 4th ed. Sussex: John Wiley & Sons, Ltd; 2005.
2. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005;58(4):323–37.
3. Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. Am J Epidemiol. 1996;143(11):1165–73.
4. Wessinger S, Kaplan M, Choi L, Williams M, Lau C, Sharp L, et al. Increased use of selective serotonin reuptake inhibitors in patients admitted with gastrointestinal haemorrhage: a multicentre retrospective analysis. Aliment Pharmacol Ther. 2006;23(7):937–44.
5. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. J Am Med Assoc. 1995;274(22):1800–4.
6. Ryan PB, Madigan D, Stang PE, Marc Overhage J, Racoosin JA, Hartzema AG, et al. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012;31(30):4401–15.
7. Stang PE, Ryan PB, Dusetzina SB, Hartzema AG, Reich C, Overhage JM, et al. Health outcomes of interest in Observational Data: issues in identifying definitions in the literature. Health Outcome Res Med. 2012;3(1):e37–44.
8. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug Saf (in this supplement issue). doi:10.1007/s40264-013-0097-8
9. Trifiro G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salame G, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? Pharmacoepidemiol Drug Saf 2009;18(12):1176–84.
10. Madigan DB, Ryan PB, Schuemie MJ. Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies. Ther Adv Drug Saf. 2013;4(2):53–62.
11. Ryan PB, Stang PE, Overhage JM, Suchard MA, Hartzema AG, DuMouchel W, et al. A comparison of the empirical performance of methods for a risk identification system. Drug Saf (in this supplement issue). doi:10.1007/s40264-013-0108-9
12. Reich CG, Ryan PB, Schuemie MJ. Alternative outcome definitions and their effect on the performance of methods for observational outcome studies. Drug Saf (in this supplement issue). doi:10.1007/s40264-013-0111-1
13. Gould AL. Accounting for multiplicity in the evaluation of "signals" obtained by data mining from spontaneous report

adverse event databases. Biometrical J Biometrische Zeitschrift. 2007;49(1):151–65.
14. Schuemie MJ, Madigan D, Ryan PB. Empirical performance of Longitudinal Gamma Poisson Shrinker (LGPS) and Longitudinal Evaluation of Observational Profiles of Adverse events Related to Drugs (LEOPARD): lessons for developing a risk identification and analysis system. Drug Saf (in this supplement issue). doi:10.1007/s40264-013-0107-x
15. Norén GN, Bergvall T, Ryan PB, Juhlin K, Schuemie MJ, Madigan D. Empirical performance of the calibrated self-controlled cohort analysis within temporal pattern discovery: lessons for developing a risk identification and analysis system. Drug Saf (in this supplement issue). doi:10.1007/s40264-013-0095-x
16. Madigan D, Schuemie MJ, Ryan PB. Empirical performance of the case-control method: lessons for developing a risk identification and analysis system. Drug Saf (in this supplement issue). doi: 10.1007/s40264-013-0105-z
17. Ryan PB, Schuemie MJ, Gruber S, Zorych I, Madigan D. Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. Drug Saf (in this supplement issue). doi:10.1007/s40264-013-0099-6
18. Suchard MA, Zorych I, Simpson SE, Schuemie MJ, Ryan PB, Madigan D. Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system. Drug Saf (in this supplement issue). doi:10.1007/s40264-013-0100-4
19. Ryan PB, Schuemie MJ, Madigan D. Empirical performance of the self-controlled cohort method: lessons for developing a risk identification and analysis system. Drug Saf (in this supplement issue). doi:10.1007/s40264-013-0101-3
20. DuMouchel W, Ryan PB, Schuemie MJ, Madigan D. Evaluation of disproportionality safety signaling applied to health care databases. Drug Saf (in this supplement issue). doi:10.1007/s40264-013-0106-y
21. Hernandez-Diaz S, Varas-Lorenzo C, Garcia Rodriguez LA. Nonsteroidal antiinflammatory drugs and the risk of acute myocardial infarction. Basic Clin Pharmacol Toxicol. 2006;98(3):266–74.
22. Loke YK, Kwok CS, Singh S. Comparative cardiovascular effects of thiazolidinediones: systematic review and meta-analysis of observational studies. BMJ. 2011;342:d1309.
23. Green J, Czanner G, Reeves G, Watson J, Wise L, Beral V. Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort. BMJ. 2010;341:c4444.
24. Cardwell CR, Abnet CC, Cantwell MM, Murray LJ. Exposure to oral bisphosphonates and risk of esophageal cancer. J Am Med Assoc. 2010;304(6):657–63.
25. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 1). EMA/95098/2010 (cited 2013 January 23). Available from: http://www.encepp.eu/standards_and_guidances/documents/ENCePPGuideofMethStandardsinPE.pdf.
26. Gagne JJ, Nelson JC, Fireman B, Seeger JD, Toh D, Gerhard T, et al. Taxonomy for monitoring methods within a medical product safety surveillance system: year two report of the Mini-Sentinel Taxonomy Project Workgroup (Workgroup) 2012 (cited 2012 October 29). Available from: http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Taxonomy-Year-2-Report.pdf.
27. Vessey MP, Doll R. Investigation of relation between use of oral contraceptives and thromboembolic disease. Br Med J. 1968;2(5599):199–205.
28. Vessey M. Learning how to control biases in studies to identify adverse effects of drugs. J R Soc Med. 2007;100(11):526–7.
29. Jick H, Vessey MP. Case-control studies in the evaluation of drug-induced illness. Am J Epidemiol. 1978;107(1):1–7.
30. Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008;19(6):766–79.
31. Perrio M, Waller PC, Shakir SA. An analysis of the exclusion criteria used in observational pharmacoepidemiological studies. Pharmacoepidemiol Drug saf. 2007;16(3):329–36.