THE EUROPEAN JOURNAL OF CONTRACEPTION & REPRODUCTIVE HEALTH CARE, 2017
VOL. 22, NO. 4, 250–259
https://doi.org/10.1080/13625187.2017.1361528

Taylor & Francis
Taylor & Francis Group

REVIEW

Real-world research and the role of observational data in the field of gynaecology – a practical review

Oskari Heikinheimo[a], Johannes Bitzer[b] and Luis García Rodríguez[c]

[a]Department of Obstetrics and Gynaecology, University of Helsinki and Kätilöopisto Hospital, University, Helsinki, Finland; [b]Department of Obstetrics and Gynaecology, University Hospital, Basel, Switzerland; [c]Spanish Centre for Pharmacoepidemiological Research, Madrid, Spain

ABSTRACT

Objectives: In the context of women's health, we examine (1) the role that observational ('real-world') studies have in overcoming limitations of randomised clinical trials, (2) the relative advantages and disadvantages of different study designs, (3) the importance of outcome data from observational studies when making health-economic or clinical decisions, and (4) provide insights into changing perceptions of observational clinical data.

Methods: PubMed and internet searches were used to identify (i) guidance and expert commentary on designing, conducting, analysing, and reporting clinical trials or observational studies, (ii) supporting evidence of the rapid growth of observational ('real world') studies and publications since the turn of millennium in the fields of contraception, reproductive health, obstetrics or gynaecology.

Results: The rapidly growing use and validation of large, computerised medical records and related databases (e.g., health insurance or national registries) have played a major part in changing perceptions of observational data among researchers and clinicians. In the past 10 years, a distinct increase in the number of observational studies published tends to confirm their growing acceptance, appreciation and use.

Conclusions: Observational studies can provide information that is impossible or infeasible to obtain otherwise (e.g., impractical, very expensive, or ethically unacceptable). Greater understanding, dissemination, uptake and use of observational data might be expected to drive ongoing evolution of research, data collection, analysis, and validation, in turn improving quality and therefore credibility, utility, and further application by clinicians.

ARTICLE HISTORY
Received 27 April 2017
Revised 21 July 2017
Accepted 25 July 2017

KEYWORDS
Clinical trials; epidemiology; evidence-based practice; observational studies; review; real-world data; women's health

Introduction

Aims

This review aims to examine the role that observational (real-world) studies can and do play in overcoming limitations associated with other study designs including randomised clinical trials, and discusses the relative advantages and disadvantages of different study designs in the field of gynaecology. It also seeks to understand improvements in both the quality and perceptions of observational versus randomised clinical data, as well as the potential importance of considering outcome data from observational studies when making health-economic or clinical decisions. Furthermore, observational data can be useful when developing guidelines and protocols to inform clinical practice, particularly in the context of gynaecology and reproductive health care. Improvements in study methodology and the completeness of data acquisition, along with the importance of validating data, are also addressed, since historically the quality of observational studies has varied hugely. Randomised clinical trials have also had similar problems in the initial generation of their use, especially in developing countries.

What are observational data?

Observational data can be defined as data generated from experience with routine medical care that has been systematically recorded, originally as administrative claims (e.g., insurance/payers' administration), in electronic medical records (e.g., clinical management in primary/secondary care databases) or national registries (e.g., birth or cancer registries), in a manner that can be used for the purposes of research [1]. The data should be derived from heterogeneous, large, (ideally) unselected populations. An important factor when considering the use of observational data is that acceptable quality standards for the data must have been established for the source(s) from which the data originated [2].

Comparison of clinical trial and observational study designs

Several trial designs exist (Figure 1). It is important to understand the relative strengths and weaknesses of these clinical trial and observational study types (Table 1). However, a clear-cut distinction does not necessarily exist between pragmatic trials and experimental (clinical) trials;

CONTACT Oskari Heikinheimo oskari.heikinheimo@helsinki.fi Department of Obstetrics and Gynaecology, University of Helsinki and Kätilöopisto Hospital, Helsinki University Central Hospital, PO Box 610, 00029 HUS, Helsinki, Finland

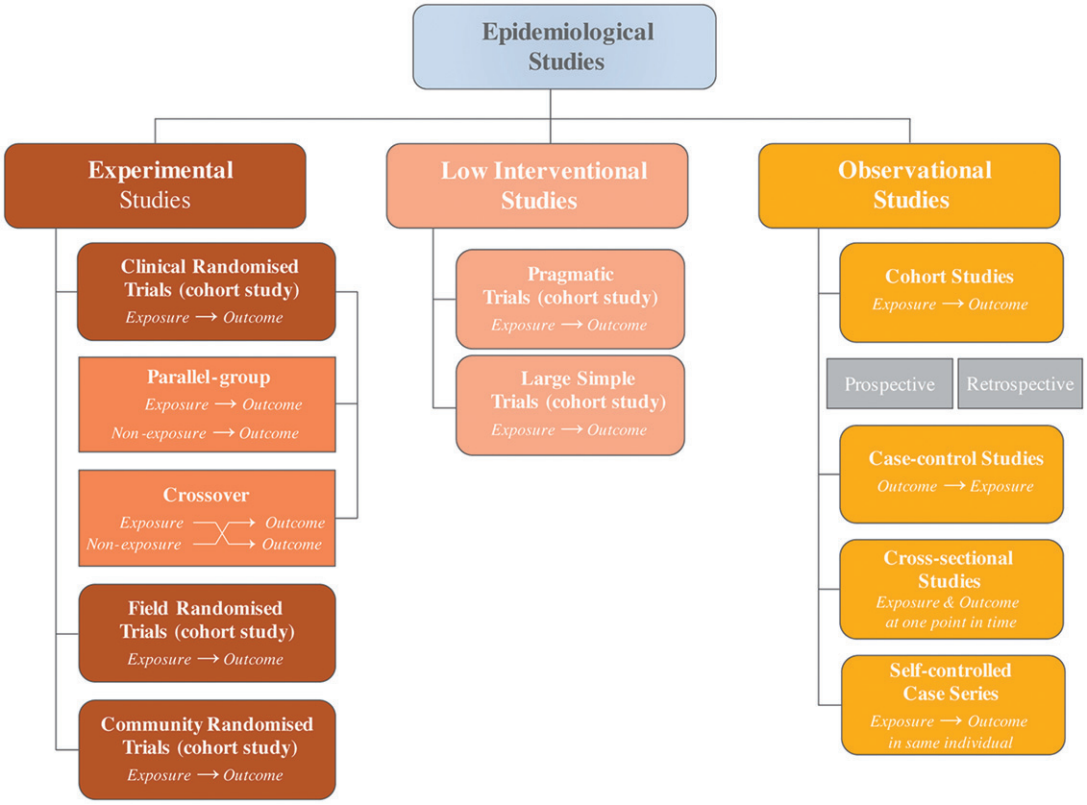

Figure 1. Observational study and randomised clinical trial (experimental) designs.

most clinical trials are neither entirely pragmatic nor entirely experimental – they lie on a continuum [3–8].

Changing perceptions of observational studies versus randomised clinical data

Data derived from randomised clinical trials are often held in higher regard than those from observational studies, with a perception that observational studies are less valid and can inflate positive treatment results or underestimate safety data as compared with randomised clinical trials [9]. However, comparative analyses of observational and clinical study results have demonstrated that data from well designed, observational studies do not consistently or systematically overestimate the magnitude of treatment effects when compared with data from randomised clinical trials on the same topic [9,10].

Evidence levels

'Good' and 'bad' examples of prospective clinical trials and observational studies exist, depending on a range of variables. For this reason, systematic ratings of evidence levels have been developed, such as that of the centre for evidence-based medicine (CEBM), which rates evidence from level 1 to 5 (best to worst, with detailed criteria). Recommendations are then graded A–D dependent on the level of evidence used (Table 2) [11]. This system ensures consistent verification of data quality and credibility; however, it does have its flaws. It is important to note that just because a study is a randomised clinical trial does not necessarily mean that it should be ranked as evidence level 1; it may be very poorly designed and/or implemented. Furthermore, based on current rating systems, such as the one illustrated in Table 2, observational data will never offer level 1 evidence. Whether this remains appropriate is a matter of debate given the increasing importance and credibility of observational data. In this context, it is noteworthy that recommendations graded below A are less likely to be adopted in either clinical practice or clinical practice guidelines, which means that observational data cannot be effectively deployed at present.

Rationale for growing significance of observational study data

Ideally, evidence-based good clinical practice relies on a combination of clinical experience (both personal and published real-world data) and experimental (clinical) research implemented, for example, via clinical practice guidelines (Figure 2).

As experience is accumulated over time, our ability to detect treatment effects that exist increases. Consequently, as more emphasis is placed on effectiveness in specific populations and health care systems, observational data become increasingly relevant. Furthermore, methodological improvements, including more sophisticated choice of datasets, better statistical methods and development of guidance supporting consistency of approach, all mean that observational studies are becoming a much more credible study genre than before – data quality is essential [9,12,13].

Methods are being actively developed and implemented in an attempt to further strengthen observational data collection and analysis. Previously (up to approximately the 1990s), many observational studies involved direct contact with participants (e.g., field-based studies) to obtain information. Since then, computerised sources of information (electronic databases) have gradually come to dominate the field. The evolution of large, computerised health

Table 1. Strengths and weaknesses of different types of randomised clinical trials and observational studies [3,5–8].

Study type	Strengths	Weaknesses
Randomised clinical trials		
Parallel group: Participants randomly allocated to either treatment/intervention or control/placebo groups for study duration *Best for:* Investigating the causal effect of an intervention	• Distribution of confounders is unbiased • Potential concealment of allocation (blinding) from participants and clinicians helps to reduce selection, allocation and outcome (performance) biases • Statistical analysis is facilitated • Complete baseline information (e.g., body-mass index, smoking status, etc.) can be obtained	• Costly in terms of time and money • Potential selection bias (e.g., healthy volunteer effect – participants are healthier than general population, leading to lower outcome rates than expected) • Hawthorne effect (an increase in outcome under study in participants who are aware of being observed) • Possibility of ethical problems (e.g., using a placebo group as a control when testing treatment for a life-threatening condition such as haemorrhage) • Restricted population size • Misclassification of exposure, particularly in long-term trials • Difficult or impossible in minors
Crossover: Participants randomly allocated to either treatment/intervention or control/placebo groups, but then switched between groups at a predetermined time point during the study	• Participants act as own controls; reduced sample size required due to statistical advantage • All participants receive treatment at least some of the time • Maintenance of blinding (benefits of blinding and best application are as described above)	• Relevant only if outcome (e.g., symptoms) is reversible over time • Placebo or alternative treatment received by all patients at some point • Prolonged or unknown washout period (time required for elimination of drug and/or its metabolites from the bodies of participants) • Unsuitable for treatments with permanent effects
Field-randomised: Similar to either parallel-group or crossover randomised clinical trials (not necessarily requiring placebo control), but generally involve participants who are living at home in their normal environment, rather than being 'captive' in hospitals or outpatient clinics [6] *Best for:* Investigating preventive measures, such as immunisations or health education	• Less stringent inclusion/exclusion criteria allow wider recruitment pool • Results are more generalisable to broader population (external validity) than those from 'clinical' randomised trials	• Less stringent inclusion/exclusion criteria can make data analysis complex
Community-randomised: Participants from specific geographic communities or 'clusters' (e.g., caregivers, hospitals, residential homes, etc.) randomised to either treatment/intervention or control arms; generally characterised by small numbers of clusters with a large number of patients in each cluster [5] *Best for:* Evaluating delivery of health care services, effects of educational interventions, or effects of organisational changes.	• Useful for comparative effectiveness research – focused on understanding effects of an intervention in a routine clinical setting • Pragmatic method of measuring effectiveness of an intervention on a large scale • Useful when randomisation at individual level is inappropriate or impossible	• Potential complexity • Intracluster correlation requires greater patient numbers than in individual randomised clinical trials • Potential for selection bias (patients may not have access to treatments or procedures available only at other locations)
Observational ('real-world'), prospective or retrospective studies[a] Cohort: No allocation of treatment exposure made by researcher *Best for*: Investigating the absolute effect of risk factors on outcomes	• Ethically sound • Subject matching possible • Outcome-assessment standardisation possible • Administration simpler and cheaper than randomised clinical trial • Allows inclusion of minors	• No randomisation • Confounders • Blinding is problematic/cannot be done • Prolonged follow-up or large sample sizes required for rare outcomes • Complete background information (e.g., body-mass index, smoking status, etc.) often difficult to obtain/manage; such factors may be a source of bias (*Note that the weaknesses above, for a cohort study, apply to any observational study to a greater or lesser extent*)
Cross-sectional: Investigates relationship between disease (or other health-related characteristic) and other variable of interest in defined population at same time – exposure and outcomes determined simultaneously *Best for:* Quantifying prevalence of a disease/risk factor, or diagnostic test accuracy	• Low cost • Ethically sound	• No randomisation • Cannot confirm temporal association • Susceptible to recall bias • Confounders • Potential for imbalanced group sizes
Case-control: Involves selection of patients with certain outcome/disease, and matched controls lacking that outcome/disease, followed by analysis of whether exposure of subjects to an investigative factor has occurred; outcome/disease is fixed but exposure occurs at random [7] Nested case-control: Variant of the case-control study adopted for its cost-effective and objective sampling strategy [8]	• Low cost and rapid • Ethically sound	• No randomisation • Exposure status relies on recall in interview-based studies • Confounders Potential bias in selection of cases and controls
Self-controlled case series: A case-only method in which individuals act as their own controls [7] *Best for:* Acute events and transient exposures for which periods of exposure risk can be clearly defined	• No separate controls are required • Automatic control of any fixed confounders	• Exposure probability not necessarily affected by the prior occurrence of an outcome • Observation period may not be independent of timing of events

Adapted from Centre for Evidence-based Medicine (CEBM) [3].

[a]Prospective studies investigate outcomes (e.g., disease development) during a study period and relate this to other factors (e.g., suspected risk or protective factors); retrospective studies do the reverse, by investigating suspected risk or protection-factor exposure in relation to a predetermined outcome at the study outset. For example, in a prospective cohort study, the incidence or prevalence of the outcome of interest in the subjects is unknown at the time that investigators begin identifying subjects and collecting exposure information, whereas in a retrospective cohort study, the outcomes have already occurred at the time that investigators plan the study and begin identifying subjects.

databases has contributed to improving the perception, uptake and application (comparative effectiveness research) of observational data. The ability to analyse/query such databases in real time is particularly relevant because rapid responses are increasingly important in managing a wide variety of public health issues, such as risk-benefit of medical or surgical interventions [12–18].

The influence of observational studies has grown in recent years

It is notable that the number of observational studies published has increased markedly in the past 10 years (data not shown), indicating their growing acceptance and appreciation, and perhaps recognising their value in providing data complementing that from randomised clinical trials. It may also suggest increasing observational research in those fields. Indeed, it is now understood that observational studies, including various data sources (e.g., disease/hospital registries, primary/secondary care medical records, databases of health-insurance claims, and surveys), can provide information that is impossible or infeasible to obtain otherwise (i.e., impractical, too expensive, or ethically unacceptable) [19].

Table 3 presents circumstances under which observational studies and real-world data sources are valuable and potentially supportive of data from randomised clinical trials [20], along with examples of key trials in the fields of contraception and reproductive health care, and the impact that they have had on clinical practice [21–36]. The first column covers particular requirements that an observational study might address; the second column provides specific examples of studies meeting those requirements; the third column provides insight into how data meeting the requirements might impact on clinical practice, i.e., additional context.

Table 2. CEBM grades of recommendation.

Grade	Basis of recommendation
A	Level 1 studies (e.g., homogeneous[a] systematic reviews of randomised clinical trials or prospective cohort studies; or individual randomised clinical trial with narrow confidence intervals; or CDR validated in different populations)
B	Level 2 (e.g., homogeneous systematic reviews of retrospective cohort studies or untreated control groups in randomised clinical trials; or individual cohort study, including low-quality randomised clinical trial) or Level 3 studies (e.g., homogeneous systematic reviews of case-control studies; or individual case-control studies) or extrapolations from level 1 studies
C	Level 4 studies (e.g., case series, poor quality cohort and case-control studies; superseded or non-independent reference standards; lack of sensitivity analysis) or extrapolations from levels 2 or 3 studies
D	Level 5 evidence (e.g., expert opinions without critical appraisal, based on physiology, basic bench research, or 'first principles') or troublingly inconsistent or inconclusive studies of any level

Adapted from CEBM [11].
[a]Homogeneity indicates freedom from worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. CDR, clinical decision rule (algorithm or scoring system leading to a prognostic decision or diagnostic category).

Limitations and strengths of observational data

Limitations of randomised clinical trials versus observational studies

The strengths and weaknesses of the basic clinical and observational study types presented in Table 1 may be condensed to illustrate the strengths of observational studies versus randomised clinical trials as a whole (Table 4).

In general, data collected from patients in real-world settings reflect what physicians see in routine clinical practice (e.g., adherence patterns and training needs), improve understanding of infrequent events, and enable analysis of long-term risks and benefits of given interventions [20].

Inherent limitations of observational studies

Bias and confounders, and their management are vital considerations in all study types. The most important biases are generally those produced in (i) the definition and selection of the study population, and (ii) data collection. Confounders are those variables that have an impact on the measured outcome independent from the intervention under investigation; in other words, the association between different determinants of an effect (outcome) in

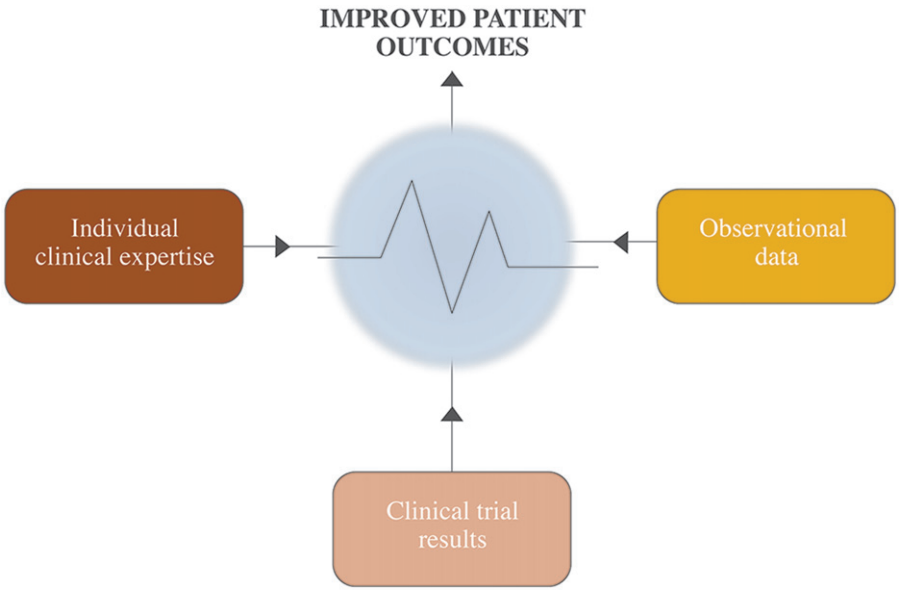

Figure 2. The evidence-based medicine triad.

Table 3. Circumstances under which observational studies are valuable and can be supportive of randomised clinical trials.

Circumstance	Study examples relevant to contraception and reproductive health care	Impact on clinical practice integrated into guidelines
When large studies are needed to ascertain an outcome (e.g., to assess infrequent or long-term effects)	• Royal College of General Practitioners' Oral Contraception Study – *Cohort study* [21,22] **Objective:** To determine whether the mortality risk among women who have used oral contraceptives differs from that of never users • Nurses' Health Study – *Cohort study* [23] **Objective:** To determine whether use of oral contraceptives is associated with all-cause and cause-specific mortality • Etude Epidémiologique auprès de femmes de l'Education Nationale (E3N) study – *Cohort study* [24,25] **Objective:** To investigate risk factors for cancer in women • Risk factors for venous thromboembolism in 1.3 million pregnancies: a nationwide prospective cohort – *Cohort study* [26] **Objective:** To quantify risk factors for venous thrombo-embolism during pregnancy and the puerperal period • Impact of estrogen type on cardiovascular safety of combined oral contraceptives – *Cohort study* [27] **Objective:** To assess the risks of serious cardiovascular events (including VTE, PE, ATE, AMI, CVA) associated with short- and long-term use of DNG/EV, oCOC and LNG in a study population representative of actual users of the individual preparations	Provides important information about benefit-risk evaluation for the health practitioner
In instances when treatment adherence might have an impact on outcome	• Prescribing of cyproterone acetate/ethinylestradiol in UK general practice: a retrospective descriptive study using The Health Improvement Network – *Descriptive study* [28] **Objective:** To investigate prescribing patterns of cyproterone acetate/ethinylestradiol in the UK before and after the 2013 prescribing guidance • The continuation rates of long-acting reversible contraceptives in UK general practice using data from The Health Improvement Network – *Cohort study* [29] **Objective:** To determine the continuation rates of new users of long-acting reversible contraceptive (LARC) methods in the UK, using data from general practice • Age, parity, history of abortion and contraceptive choices affect the risk of repeat abortion – *Cohort study* [30] **Objective:** To determine whether use of an IUD/IUS post abortion is more effective than user-dependent contraceptive methods in reducing the need of further induced abortion	Implementation research is very important for obtaining information about the behaviour of the target group in relation to the outcome
When training of providers might have an impact on outcome	• Which factors are associated with trainees' confidence in performing obstetric and gynaecological ultrasound examinations? – *Cross-sectional study* [31] **Objective:** To explore the association between clinical training characteristics and trainees' level of confidence in performing ultrasound scans independently	Adaptation of programmes/protocols in clinical practice
In instances when a timely result is needed	• Maternal vaccination against H1N1 influenza and offspring mortality: population-based cohort study and sibling design – *Cohort study* [32] **Objective:** To investigate the safety of H1N1 influenza vaccination during pregnancy • Oral contraceptives, venous thrombosis, and varicose veins. Royal College of General Practitioners' Oral Contraception Study – *Cross-sectional study* [33] **Objective:** To assess the relationship between oral contra ceptive use and risk of idiopathic deep and superficial thrombosis in the leg	Important in situations in which it would be politically or ethically unacceptable to deny access to an intervention
When multiple treatment solutions are available	• Population-based evaluation of the effectiveness of two regimens for emergency contraception – *Cohort study* [34] **Objective:** To estimate and compare the effectiveness of the levonorgestrel and Yuzpe regimens for hormonal emergency contraception in routine clinical practice • Endometriosis and its treatment with danazol or lupron in relation to ovarian cancer – *Case-control study* [35] **Objective:** To test whether exogenous androgens used for endometriosis may be associated with ovarian cancer	Impact on guidelines regarding the use of emergency contraception, and androgen use in endometriosis
When wanting to explore population subsets	• Danish Sex Hormone Register Study: association of hormonal contraception with depression – *Cohort study* [36] **Objective:** To determine whether hormonal contraception is positively associated with subsequent use of antidepressants and a diagnosis of depression at a psychiatric hospital	Providing associations and hypotheses, which have to be explored further

Adapted and expanded from Dreyer et al. [20]. AMI: acute myocardial infarction; ATE: acute thromboembolism; CVA: cerebrovascular accidents; EV: estradiol valerate; DNG: dienogest; IUD: intrauterine device; IUS: intrauterine system; LNG: levonorgestrel; oCOC: other combined oral contraceptive; PE: pulmonary embolism; VTE: venous thromboembolism.

Table 4. Limitations of randomised clinical trials versus observational studies.

Randomised clinical trials	Observational studies
Expensive	Lower costs in regulatory and administrative terms
Can be time consuming	Usually less time consuming (important when public safety is at stake, and regulatory actions must be taken)
Strict exclusion and inclusion criteria may cause study population to differ from target population	Representative of target population
Treatment efficacy is assessed, involving strict treatment strategies and/or patient monitoring that may not be practical in a real-world setting	Treatment effectiveness under real-world conditions (routine clinical practice) is assessed
Potentially unethical (e.g., teratogenic effects)	Avoids ethical issues in e.g., the study of noxious substances
Achieving adequate sample sizes for statistical power can be difficult	Large sample sizes[a] ensure statistical power (the large sample has to be selected or weighted to represent the population from which sample is drawn)

[a]While large sample size can be a strength of observational studies, it does carry a risk of a given outcome being 'over-dramatised', particularly if that outcome is negative, such as death.

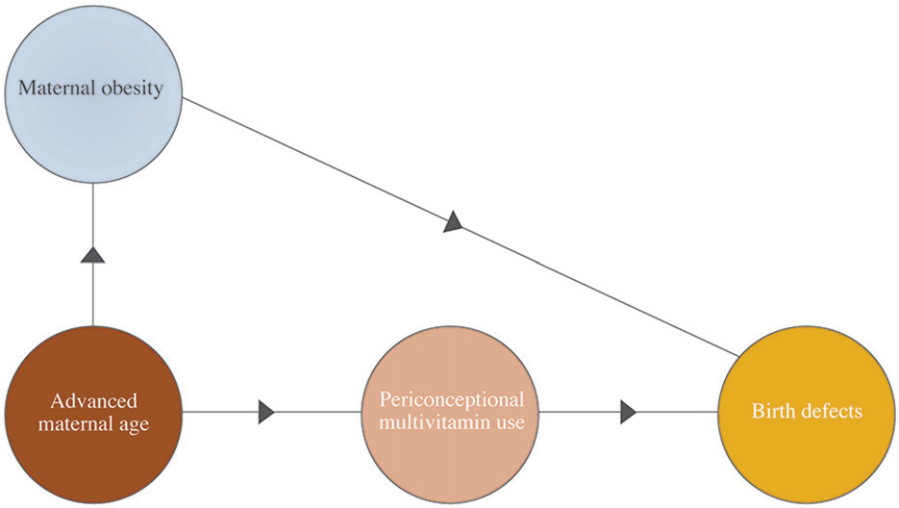

Figure 3. Causal diagram of factors confounding the relationship between maternal obesity and birth defects. Adapted from Hernán et al. [38].

the population [37]. Careful study design and analysis are needed to avoid bias and adjust for confounders.

Confounders, and differences between individuals exposed and unexposed to treatment

An investigator is not necessarily directly interested in confounders but they may affect the results of a study if not taken into account, and can contribute to an observational study receiving a poor evidence rating (see Table 2) [11]. Statistical analyses typically involve three variables: exposure, outcome, and confounders. The causal question being investigated usually determines exposure and outcome but confounders can be unclear; they require identification, and analyses should be adjusted accordingly. An example of a causal diagram (Figure 3) illustrates how confounders can play a role in the epidemiology of birth defects. Advanced maternal age may increase the risk of maternal obesity, which is associated with certain birth defects. However, advanced maternal age also increases the chance of periconceptional multivitamin use (folate supplementation may reduce the risk of neural tube defects) [38].

Confounding by indication

Patients in routine clinical practice who receive drugs, or see a gynaecologist, are quite different from those who do not, whereas patients in the two arms of a randomised clinical trial are to all intents and purposes equal except for the intervention. Confounding by indication is a generalised problem in the real-world setting; it occurs when a variable is a risk factor for a condition among non-exposed individuals and, at the same time, is the reason to receive treatment without being an intermediate step in the causal pathway between the treatment and the condition [39].

It is possible to use the association of hormonal contraception with depression as an example of how steps can be taken in an effort to avoid potential confounding by indication. Pregnant women (having no use for hormonal contraception and a potentially cautious attitude towards antidepressant use) as well as those with previous mental diseases might be excluded from a prospective observational study. Risk factors for being treated for depression can include help-seeking behaviour, contact with the health care system, attitudes of health care professionals and so forth. A statistical approach, such as an analysis comparing each woman with herself for antidepressant use pre- and post-hormonal contraception period, would facilitate elimination of many possible confounders considered to be relatively constant over time (e.g., alcohol, smoking habits, body mass index, education, attitudes for use of medicine in general, low social status, etc.).

One approach to reduce confounding by indication is to compare drugs with similar indications/contraindications, or to have access to detailed information on treatment indications that allow adjustments in the analysis.

Incomplete data on predictors

Adjustment for confounding factors is mandatory when analysing observational data. All relevant factors must be

available, and such data should be confirmed as being as valid and complete as possible. In this respect, databases of health insurance claims are more prone to selection bias than electronic medical records; e.g., in the USA, patients with health insurance are automatically likely to be from a higher income bracket and therefore to have received better health care and have a better standard of living than patients without health insurance.

Incomplete data on events

Identification (and full record keeping) of all relevant events should be ensured, as well as that all identified events are actually relevant. Validation studies are key here to ensure both the sensitivity and specificity of an investigational strategy.

Selection procedure

This is particularly important in (but not unique to) case-control studies. Controls should be representative of the population that gave rise to the cases, and must be sampled independently of exposure. It is also important that exposure status is not influenced by individual responses to a drug; selection of prevalent users should be avoided. However, new users should receive long-term follow-up, if relevant to the aims of the study. The French E3N study (Table 3) represents a good example of such follow-up, in which participating women have received questionnaires every two years, starting in 1990; the study is ongoing to date [24,25].

Key indicators of high-quality observational data

Reproducibility, data completeness and predictors

It is essential that evidence obtained from a given data source can be replicated, or is comparable with data from another source (e.g., randomised clinical trials or other observational studies, disease registries or surveys on the same topic). Data completeness is highly important. Data for as many predictors as possible for the condition being studied should be available; such predictors include those of *characterisation* (e.g., comorbidity, prior drug use, lifestyle factors such as smoking or alcohol use, demographics such as age, gender, body-mass index and deprivation), and those of *outcome* (identification of study events using clear and reproducible strategies based on a combination of diagnostic codes, free text, and laboratory test values – as applicable).

Validation of data

Data validation (the verification of reproducibility or comparability with other studies as described above) is commonly done by comparing data recorded in a computerised database with that on paper medical records, or on other databases (e.g., comparison of hospitalisations recorded in a primary care database with those recorded in a secondary care database). This is also the key to confirming the validity of methodology used to identify the relationship between exposure (if possible, for example, prescription issue should be related to prescription filling or purchase to confirm exposure) and study events (outcomes). The methodological steps of the validation process should be reported, clear and systematic, and include description of assumptions and inferences, as well as interpretation of potential under- or over-estimation of the outcome of interest. An example of such validation is that for an observational study of uterine fibroids in the UK, based on data in The Health Improvement Network (THIN), and reported by Martín-Merino et al. [40].

It should be noted that no uniform criteria exist for validating data in a database, although various approaches have been detailed, e.g., for diagnostic coding within the general practice research database (GPRD) [41], for database selection and use in pharmaco-epidemiological research [42], and for quality assessment of health care data used in clinical research [43].

Ascertainment and case validation when using a primary care database in observational studies

Step 1 – Choosing a primary care database

The GPRD and THIN are examples of large primary care databases in the UK that have acceptable internal and external validities overall, supported by peer-reviewed studies by external researchers as well as data provided by the database owners. As such, they represent sound 'starting-point' databases for an observational study [44]. Nonetheless, it should be remembered that even such databases of generally acceptable (and accepted) validity are not automatically valid for all the conditions and diagnoses for which a researcher might wish to use them.

Step 2 – Conducting a computerised search

Having chosen a database with acceptable validity as a first step, care needs to be taken in the second step of conducting an initial, computerised search of that database. Factors to consider include (i) establishing a robust definition of outcome of interest by implementing specific diagnostic algorithms based on codes listed in an appropriate clinical dictionary, and (ii) making use of specific, objective, eligibility criteria [44].

Step 3 – Manually reviewing computerised patient profiles (ascertainment)

It is vital to assess whether the validity of results from the initial, computerised search is acceptable (i.e., a confirmation rate close to 90%) or whether more information needs to be obtained. This requires conscientious manual review of the computer profiles of patients identified in the search, including (i) information stored as free text (e.g., physician narratives, diagnostic procedures, referral/discharge letters), and (ii) assignment of case status (i.e., probable, possible, or non-case) for each patient [44].

Free text can conceal substantial additional information from the computerised search regarding diagnoses. Manual review of free text potentially allows 'read code' misclassifications to be identified, and accurate dating of diagnoses/referrals [45].

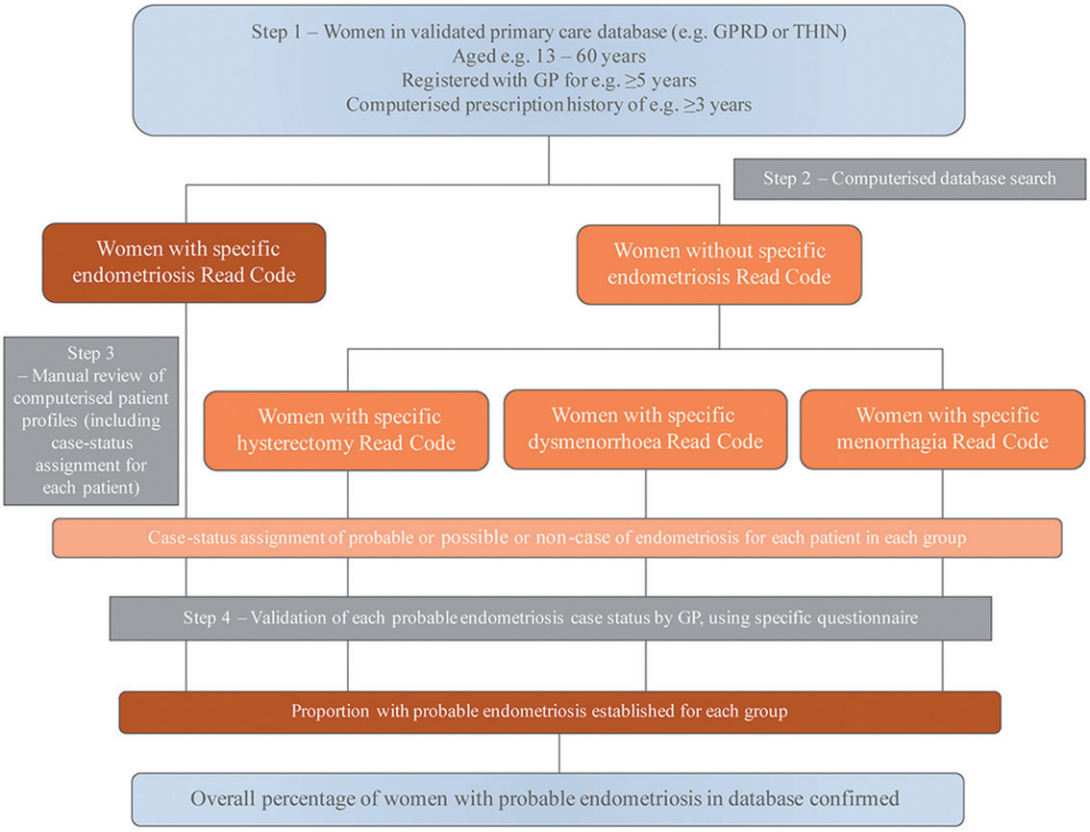

Figure 4. Flowchart exemplifying typical steps for ascertainment and case validation when using a primary care database in an observational study. GP: general practitioner; GPRD: General Practice Research Database; THIN: The Health Improvement Network.

Step 4 – Case status validation by general practitioner (GP)

The assignment of case status in the third step requires validation by the relevant GP for each case. This may be done by sending specifically designed questionnaires to collaborating GPs; in addition, the database owner may be requested to provide anonymised copies of original medical records (e.g., consultant letters, post-mortem reports). Time and expense can be saved by doing such validation on a random sample of the patient records identified in the database, provided that a high positive predictive value is anticipated on the basis of previous ascertainment [44]. Figure 4 illustrates steps 1–4 described above.

What guidelines are available for conducting, interpreting and reporting observational studies?

Clinicians, regulators, patients, payers and policy makers will only take observational data into account if the quality is assured. A number of standards and guidelines have been published to support the conduct, analysis and reporting of observational studies and data (see Panel) [2,46–52].

When reading an article reporting results from a clinical trial or observational study, critical appraisal of the information is essential to ensure that the information and data presented are of good quality and fit for purpose. Reporting should be transparent, allowing the reader a thorough and unambiguous understanding of the research (i.e., what was done, what was found, and what conclusions were made), including assessment of the strengths and weaknesses in study design, conduct, and analysis. The STROBE guidance represents a good example of how to approach such assessment [52].

Panel. Standards and guidelines for observational studies.

Study design

- Agency for Health care Research Quality (AHRQ): Developing a protocol for observational comparative effectiveness research [46]
- European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) checklist for study protocols [47]
- ENCePP Guide on methodological standards in pharmacoepidemiology (Revision 5) [48]
- International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Good research practices for retrospective database analysis task force report (Parts I–III) [49–51]

Data interpretation

- The GRACE checklist: A validated assessment tool for high quality observational studies of comparative effectiveness [2]

Data reporting

- The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies [52]

Conclusions

Robust evidence-based practice is essential for ensuring the current and future quality of health care. Owing to the respective strengths and weaknesses inherent to clinical trials and observational studies (involving routine clinical experience), it is highly important to consider a combination of study types when making clinical decisions and, as far as possible, not be reliant on just one type or the other. No study setting is a 'Holy Grail' and data derived from different study settings are often complementary.

Clinical decision-making, including the development of guidelines and protocols for patients or clinicians, can be distilled to a systematic approach involving the 'five As': **Ask** the correct question (this might arise due to a recurring problem in the clinic, be raised during a conference, through interaction with colleagues, or while reading journal articles); **Acquire** information (clinical trial and/or observational study and/or systematic review results) to address the question; **Appraise** the information in a critical and systematic manner (can the data be validated? What level/grade of evidence is available, i.e. are the data complete and of good quality?); **Apply** the appropriate evidence in practice (involving clinicians' consensus and, potentially, guideline or protocol development/implementation); and **Assess** (analyse) the care provided in a systematic manner to determine its true benefit (including factors such as cost-effectiveness or risk-benefit) for patients, and whether clinicians are complying with guidance/protocol(s) [53].

In recent years, observational study data have become better perceived due to the greater availability of information databases of acceptable quality, and improvements and innovations in methodology [12–18], resulting in increasing numbers of publications and a growing impact on daily practice. Greater understanding, dissemination, uptake and use of observational data might be expected to drive ongoing evolution of real-world research, data collection, analysis, and validation, thereby in turn improving quality and hence credibility, utility, and further application by clinicians.

Acknowledgements

This publication and its content are solely the responsibility of the authors, who received editorial and writing support that was paid for by Bayer AG.

Disclosure statement

The authors, Oskari Heikinheimo, Johannes Bitzer and Luis García Rodríguez, have acted as consultants to Bayer AG and received consultancy honoraria unrelated to the creation of this paper. OH is employed by the Hospital District of Helsinki and Uusimaa, the University of Helsinki and Finnish Medical Society Duodecim. He also serves occasionally on advisory boards or as invited speaker in educational events organised by Bayer Healthcare, MSD/Merck, Actavis, Gedeon Richter, and Sandoz. JB has received honoraria for lectures and participation in advisory boards from Bayer AG, Merck, Libbs, Actavis, Teva, Exeltis, Gedeon Richter, Boehringer-Ingelheim, Vifor, Lilly, Pfizer, HRA, Abbott, Mithra, Pierre Fabre, and Aspen. LAGR works for Centro Español de Investigación Farmacoepidemiológica (CEIFE), which has received research funding from Bayer AG. He has also received honoraria for serving on advisory boards for Bayer AG.

Funding

This work was supported by Bayer AG.

References

[1] Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence – what is it and what can it tell us? N Engl J Med. 2016;375:2293–2297.

[2] Dreyer NA, Bryant A, Velentgas P, et al. The GRACE checklist: a validated assessment tool for high quality observational studies of comparative effectiveness. JMCP. 2016;22:1107–1113.

[3] CEBM.net [Internet]. Oxford: Centre for Evidence-Based Medicine; [cited 2017 Feb 26]. Available from: http://www.cebm.net/study-designs/

[4] CHCUK.co.uk [Internet]. Contin: What are pragmatic clinical trials?; 2015. [cited 2017 Mar 17]. Available from: http://www.chcuk.co.uk/wp-content/uploads/2015/11/2015-11-22_Pragmatic-Clinical-Trials_CHCUK1.pdf

[5] NPCNOW.org [Internet]. Washington: National Pharmaceutical Council. Making informed decisions: assessing the strengths and weaknesses of study designs and analytic methods for comparative effectiveness research. Velengtas P, Mohr P, Messner DA, editors. 2012 February [cited 2017 Mar 19]. Available from: http://www.npcnow.org/system/files/research/download/experimental_nonexperimental_study_final.pdf

[6] OUP. Oxford; Oxford University Press. In: Smith PG, Morrow RH, Ross DA, editors. Field Trials of Health Interventions: A Toolbox. 3rd ed.; 2015 Jun 1. Chapter 1, Introduction to field trials of health interventions [cited 2017 Mar 19]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK305510/

[7] Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. BMJ. 2016;354:i4515.

[8] Cai T, Zheng Y. Evaluating prognostic accuracy of biomarkers in nested case-control studies. Biostatistics. 2012;13:89–100.

[9] Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med. 2000;342:1878–1886.

[10] Concato J. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342:1887–1892.

[11] CEBM.net [Internet]. Oxford: Centre for Evidence-Based Medicine; [cited 2017 Feb 26]. Available from: http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/

[12] Hernán MA. With great data comes great responsibility: publishing comparative effectiveness research in epidemiology. Epidemiology. 2011;22:290–291.

[13] Amarasingham R, Audet AM, Bates DW, et al. Consensus statement on electronic health predictive analytics: a guiding framework to address challenges. EGEMS (Wash DC). 2016;4:1163.

[14] Andrews EB, Margulis AV, Tennis P, et al. Opportunities and challenges in using epidemiologic methods to monitor drug safety in the era of large automated health databases. Curr Epidemiol Rep. 2014;1:194–205.

[15] Etheredge LM. A rapid-learning health system. Health Aff (Millwood). 2007;26:w107–w118.

[16] Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183:758–764.

[17] Hershman DL, Wright JD. Comparative effectiveness research in oncology methodology: observational data. J Clin Oncol. 2012;30:4215–4222.

[18] Kaggal VC, Elayavilli RK, Mehrabi S, et al. Toward a learning health-care system – knowledge delivery at the point of care empowered by big data and NLP. Biomed Inform Insights. 2016;8:13–22.

[19] Dreyer NA. Making observational studies count; shaping the future of comparative effectiveness research. Epidemiology. 2011;22:295–297.

[20] Dreyer NA, Tunis SR, Berger M, et al. Why observational studies should be among the tools used in comparative effectiveness research. Health Affairs. 2010;29:1818–1825.

[21] Hannaford PC, Iversen L, Macfarlane TV, et al. Mortality among contraceptive pill users: cohort evidence from Royal College of General Practitioners' Oral Contraception Study. BMJ.2010;340:c927

[22] Iversen L, Sivasubramaniam S, Lee AJ, et al. Lifetime cancer risk and combined oral contraceptives: the Royal College of General Practitioners' Oral Contraception Study. Am J Obstet Gynecol. 2017;216:580.e1–580.e9.

[23] Charlton M, Rich-Edwards JW, Colditz GA, et al. Oral contraceptive use and mortality after 36 years of follow-up in the Nurses' Health Study: prospective cohort study. BMJ. 2014;349:g6356.

[24] Fournier A, Berrino F, Clavel-Chapelon F. Unequal risks for breast cancer associated with different hormone replacement

therapies: results from the E3N cohort study. Breast Cancer Res Treat. 2008;107:103–111.
[25] Cadeau C, Fournier A, Mesrine S, et al. Interaction between current vitamin D supplementation and menopausal hormone therapy use on breast cancer risk: evidence from the E3N cohort. Am J Clin Nutr. 2015;102:966–973.
[26] Virkus RA, Løkkegaard E, Lidegaard Ø, et al. Risk factors for venous thromboembolism in 1.3 million pregnancies: a nationwide prospective cohort. PLoS One. 2014;9:e96495.
[27] Dinger J, Do Minh T, Heinemann K. Impact of estrogen type on cardiovascular safety of combined oral contraceptives. Contraception. 2016;94:328–339.
[28] Cea Soriano L, Asiimwe A, García Rodríguez LA. Prescribing of cyproterone acetate/ethinylestradiol in UK general practice: a retrospective descriptive study using The Health Improvement Network. Contraception. 2017;95:299–305.
[29] Cea Soriano L, Wallander MA, Andersson S, et al. The continuation rates of long-acting reversible contraceptives in UK general practice using data from The Health Improvement Network. Pharmacoepidemiol Drug Saf. 2015;24:52–58.
[30] Heikinheimo O, Gissler M, Suhonen S. Age, parity, history of abortion and contraceptive choices affect the risk of repeat abortion. Contraception. 2008;78:149–154.
[31] Tolsgaard MG, Rasmussen MB, Tappert C, et al. Which factors are associated with trainees' confidence in performing obstetric and gynecological ultrasound examinations?. Ultrasound Obstet Gynecol. 2014;43:444–451.
[32] Ludvigsson JF, Ström P, Lundholm C, et al. Maternal vaccination against H1N1 influenza and offspring mortality: population based cohort study and sibling design. BMJ. 2015;351:h5585.
[33] RCGPOCS. Oral contraceptives, venous thrombosis, and varicose veins. Royal College of General Practitioners' Oral Contraception Study. J R Coll Gen Pract. 1978;28:393–399.
[34] Leung VW, Soon JA, Lynd LD, et al. Population-based evaluation of the effectiveness of two regimens for emergency contraception. Int J Gynaecol Obstet. 2016;133:342–346.
[35] Cottreau CM, Ness RB, Modugno F, et al. Endometriosis and its treatment with danazol or lupron in relation to ovarian cancer. Clin Cancer Res. 2003;9:5142–5144.
[36] Skovlund CW, Mørch LS, Kessing LV, et al. Association of hormonal contraception with depression. JAMA Psychiatry. 2016;73:1154–1162.
[37] Delgado-Rodríguez M, Llorca J. Bias. J Epidemiol Community Health. 2004;58:635–641.
[38] Hernán MA, Hernández-Díaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002;155:176–184.
[39] Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol. 1999;149:981–983.
[40] Martín-Merino E, Wallander MA, Andersson S, et al. The reporting and diagnosis of uterine fibroids in the UK: an observational study. BMC Womens Health. 2016;16:45.
[41] Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the general practice research database: a systematic review. Br J Gen Pract. 2010;60:e128–e136.
[42] Hall GC, Sauer B, Bourke A, et al. Guidelines for good database selection and use in pharmacoepidemiology research. Pharmacoepidemiol Drug Saf. 2012;21:1–10.
[43] NIHcollaboratory.org [Internet]. Durham: National Institutes of Health Collaboratory. Assessing data quality for healthcare systems data used in clinical research (version 1.0). Zozus MN, Ed Hammond W, Green BB, et al. 2014 [cited 2017 Apr 18]. Available from: https://www.nihcollaboratory.org/Products/Assessing-data-quality_V1%200.pdf
[44] García Rodríguez LA, Ruigómez A. Case validation in research using large databases. Br J Gen Pract. 2010;60:160–161.
[45] Tate AR, Martin AG, Ali A, et al. Using free text information to explore how and when GPs code a diagnosis of ovarian cancer: an observational study using primary care records of patients with ovarian cancer. BMJ Open. 2011;1:e000025.
[46] AHRQ Publication No. 12(13)-EHC099. Rockville; Agency for Healthcare Research and Quality. In: Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, editors. Developing a protocol for observational comparative effectiveness research: a user's guide; 2013. [cited 2017 Feb 26]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK126190/
[47] ENCePP.eu [Internet]. London: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. Checklist for study protocols (revision 3). [cited 2017 Feb 26]. Available from: http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml
[48] ENCePP.eu [Internet]. London: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. Guide on methodological standards in pharmacoepidemiology (Revision 5). EMA/95098/2010. [cited 2017 Feb 26]. Available from: http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml
[49] Berger ML, Mamdani M, Atkins D, et al. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR good research practices for retrospective database analysis task force report – Part I. Value Health. 2009;12:1044–1052.
[50] Cox E, Martin BC, Van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research good research practices for retrospective database analysis task force report – Part II. Value Health. 2009;12:1053–1061.
[51] Johnson ML, Crown W, Martin BC, et al. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR good research practices for retrospective database analysis task force report – Part III. Value Health. 2009;12:1062–1073.
[52] von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344–349.
[53] de Groot M, van der Wouden JM, van Hell EA, et al. Evidence-based practice for individuals or groups: let's make a difference. Perspect Med Educ. 2013;2:216–221.