


‘Real-world’ observational studies in arrhythmia research: data sources, methodology, and interpretation. A position document from European Heart Rhythm Association (EHRA), endorsed by Heart Rhythm Society (HRS), Asia-Pacific HRS (APHRS), and Latin America HRS (LAHRS)

Christian Torp-Pedersen (Chair) ^{1*}, **Andreas Goette**²,
Peter Bronnum Nielsen¹, **Tatjana Potpara**^{3,4}, **Laurent Fauchier**⁵, **Alan John Camm**⁶,
Elena Arbelo^{7,8,9}, **Giuseppe Boriani**¹⁰, **Flemming Skjoeth**¹¹, **John Rumsfeld**¹²,
Frederick Masoudi¹³, **Yutao Guo**¹⁴, **Boyoung Joung**¹⁵, **Marwan M. Refaat**¹⁶,
Young-Hoon Kim (APHRS representative)¹⁷,
Christine M. Albert (HRS representative)¹⁸,
Jonathan Piccini (HRS representative)¹⁹, **Alvaro Avezum (LAHRS representative)**²⁰,
and **Gregory Y.H. Lip (Co-Chair)**^{21,22}

External Reviewers: Bulent Gorenek (Review co-ordinator)²³, **Nikolaos Dages**²⁴,
Francesco Violi²⁵, **Xin Du**²⁶, **Masaharu Akao**²⁷, **Eue-Keun Choi**²⁸, **Renato D. Lopes**²⁹,
Evren Emin Ozcan³⁰, **Deirdre Lane**^{21,22}, **Francisco Marin**³¹,
Christopher Peter Gale³², **Kevin Vernooy**³³, **Gulmira Kudaiberdieva**^{34,35},
Valentina Kutyla^{36,37}, **Vassil B. Traykov**³⁸, **Carolina Guevara**³⁹,
Tze-Fan Chao⁴⁰, and **Sana M. Al-Khatib**⁴¹

¹Aalborg University, Health Science and Technology, Aalborg, Denmark; ²St. Vincenz-Krankenhaus GmbH, Paderborn, Germany; ³School of Medicine, Belgrade University, Belgrade, Serbia; ⁴Cardiology Clinic, Clinical Center of Serbia, Belgrade, Serbia; ⁵Service de Cardiologie, Centre Hospitalier Universitaire Trousseau et Université de Tours, Faculté de Médecine, Tours, France; ⁶St. George's, University of London, Molecular and Clinical Sciences Research Institute, St George's, University of London, London, UK; ⁷Arrhythmia Section, Cardiology Department, Hospital Clínic, Universitat de Barcelona, Barcelona, Spain; ⁸IDIBAPS, Institut d'Investigació August Pi i Sunyer (IDIBAPS), Barcelona, Spain; ⁹Centro de Investigación Biomédica en Red de Enfermedades Cardiovasculares (CIBERCV), Madrid, Spain; ¹⁰Cardiology Division, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Policlinico di Modena, Modena, Italy; ¹¹Aalborg University, Health Science and Technology, Aalborg, Denmark; ¹²University of Colorado School of Medicine, Aurora, CO, USA; ¹³Division of Cardiology, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; ¹⁴Cardiology, Chinese PLA General Hospital, Beijing, People's Republic of China; ¹⁵Cardiology Department, Yonsei University College of Medicine, Seoul, Republic of Korea; ¹⁶Department of Internal Medicine, American University of Beirut Medical Center, Beirut, Lebanon; ¹⁷Cardiology Department, Korea University Medical Center, Seoul, Republic of Korea; ¹⁸Brigham and Women's Hospital, Boston, MA, USA; ¹⁹Duke Center for Atrial Fibrillation, Duke University Medical Center, Duke Clinical Research Institute, Durham, NC, USA; ²⁰Dante Pazzanese Institute of Cardiology, Sao Paulo, Brazil; ²¹Liverpool Centre for Cardiovascular Science, University of Liverpool, Liverpool Heart & Chest Hospital, Liverpool, UK; ²²Aalborg Thrombosis Research Unit, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; ²³Eskisehir Osmangazi University, Eskisehir, Turkey; ²⁴Department of Electrophysiology, Heart Center Leipzig, Leipzig, Germany; ²⁵Dipartimento di Medicina Interna e Specialità Mediche, Sapienza Università di Roma, Viale del Policlinico, 155 - 00162 Roma, Italy; ²⁶Department of Cardiology, Anzhen Hospital, An Ding men wai, Anzhenli, 100029 Beijing, People's Republic of China; ²⁷Department of Cardiology, National Hospital Organization Kyoto Medical Center, 1-1, Mukaihata-cho, Fukakusa, Fushimi-ku, Kyoto 612-8555, Japan; ²⁸Seoul National University College of Medicine, Cardiovascular Center, Seoul National University Hospital, 101 Daehak-ro, Chongno-gu, Seoul 03080, Republic of Korea; ²⁹Duke University Medical Center, Duke Clinical Research Institute, 200 Morris St., P.O. Box 17969, NC 27701, Durham, NC 27715, DUMC#: 3850; ³⁰Department of Cardiology,

* Corresponding author. Tel: +45 24 453 790. E-mail address: ctp@heart.dk

Published on behalf of the European Society of Cardiology. All rights reserved. © The Author(s) 2019. For permissions, please email: journals.permissions@oup.com.

Dokuz Eylul University, 35340 Balçova/İzmir, Turkey; ³¹Hospital Universitario Virgen de la Arrixaca, IMIB-Arrixaca, CIBERCV, University of Murcia, Murcia, Spain; ³²University of Leeds, LICAMM, Medical Research Council Bioinformatics Centre, LICAMM, Worsley Building, LS2 9JT Leeds, UK; ³³Department of Cardiology, Maastricht University Medical Center, PO BOX 5800, 6202 AZ Maastricht, Netherlands; ³⁴Adana, Turkey Center for Postgraduate Education and Research, Bishkek, Kyrgyzstan; ³⁵Kyrgyzstan Scientific Research Institute of Heart Surgery and Organ Transplantation, Bishkek, Kyrgyzstan; ³⁶University of Rochester Medical Center, Rochester, NY, USA; ³⁷Semmelweis University Heart Center, Budapest, Hungary; ³⁸Department of Invasive Electrophysiology, Acibadem City Clinic Tokuda Hospital, Sofia, Bulgaria; ³⁹Instituto Nacional Cardiovascular Incor, Jirón Coronel Zegarra 417, Jesús María 15072, Peru; ⁴⁰Division of Cardiology, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan; and ⁴¹Duke University Medical Center, Durham, NC 27715, USA

Received 5 July 2019; editorial decision 8 July 2019; accepted 24 July 2019; online publish-ahead-of-print 14 November 2019

Abstract

The field of observational studies or “real world studies” is in rapid development with many new techniques introduced and increased understanding of traditional methods. For this reason the current paper provides an overview of current methods with focus on new techniques. Some highlights can be emphasized: We provide an overview of sources of data for observational studies. There is an overview of sources of bias and confounding. Next There is an overview of causal inference techniques that are increasingly used. The most commonly used techniques for statistical modelling are reviewed with focus on the important distinction of risk versus prediction. The final section provides examples of common problems with reporting observational data.

Keywords

European Heart Rhythm Association • EHRA position paper • Arrhythmias • Atrial fibrillation • Epidemiological study • Real-world observational study • Registry • Nationwide cohort • Database • Methodology • Meta-analysis

Introduction

The criterion standard for demonstrating the efficacy of a clinical intervention is the randomized clinical trial (RCT). Randomization supports equal distribution of known as well as unknown confounders, and therefore, the relationship between the intervention and the outcome may be considered causal. Nevertheless, RCTs have limitations such as cost and cohort selection, and data from such trials are not available to provide evidence for the majority of clinical decisions. Most of recommendations in international cardiology guidelines are not based on randomized trials and there appears no improvement over the last 10 years.¹

For many clinical scenarios, observational data may be the highest level of evidence available.² Observational data can also be of particular use in evaluating care delivery, and effectiveness and safety of care in clinical practice. However, observational studies also carry significant limitations, especially when applied to therapeutic interventions (i.e. trying to determine effectiveness). Observational data are subject to underlying biases such as selection bias and are prone to unmeasured confounding. In an overview, 25% of observational studies were contradicted when the findings were tested in a randomized design.³ Over the last decade, there has been an exponential growth of observational data (e.g. from electronic health records, clinical registries, and other sources). This has been coupled with advances in the conduct and interpretation of observational studies to minimize these issues and guidelines/checklists have been developed for the conduct of observational studies (<https://www.strobe-statement.org>). In parallel, there is tremendous interest in utilizing observational, or ‘real world’ data to inform clinical care.

In recognizing these issues, European Heart Rhythm Association (EHRA), with additional contributions from Heart Rhythm Society (HRS), Asia Pacific Heart Rhythm Society (APHRS), and the Latin America Heart Rhythm Society (HRS) proposed a position document describing contemporary techniques for optimal conduct and

presentation of observational studies. An additional aim was to provide recommendations to encourage implementation of new designs.

This review first describes the usual data sources for observational studies, reviews common and important techniques, overviews the proper interpretation of results, and finally makes appropriate recommendations regarding the design, conduct, and interpretation of observational data. The intended reader is the clinical cardiologist that wishes to get an overview of current methodology. It is hoped that it will aid the discussion between clinicians and cardiologists. It has been attempted to cover briefly the most used current methods with focus on more recent methodology. It is a very large area that is covered, and therefore, many details are not touched in this overview.




Evidence review

This document was prepared by the Task Force with representation from EHRA, with additional contributions from HRS, APHRS, LAHRS, and CASSA, and has been peer-reviewed by official external reviewers representing all these bodies. A detailed literature review was conducted, weighing the strength of evidence for or against a specific treatment or procedure, and where data exist including estimates of expected health outcomes.

We have used a simple and user-friendly system of grading recommendations using ‘coloured hearts’ (Table 1). This EHRA grading of consensus statements does not use separate definitions of the level of evidence. This categorization, used for consensus statements, must not be considered as directly similar to that used for official society guideline recommendations, which apply a classification (Class I–III) and level of evidence (A, B, and C) to recommendations used in official guidelines.

The routine use of hearts is changed for this publication which addresses statistical methods rather than interventions. Thus, a

Table 1 Consensus statement instruction

Definitions were related to a treatment or procedure	Consensus statement instruction	Symbol
Scientific evidence that treatment or procedure is beneficial and effective. Requires at least one randomized trial or is supported by strong observational evidence and authors' consensus (as indicated by an asterisk)	'Should do this'	
General agreement and/or scientific evidence favour the usefulness/efficacy of a treatment or procedure. May be supported by randomized trials based on a small number of patients or which is not widely applicable	'May do this'	
Scientific evidence or general agreement not to use or recommend a treatment or procedure	'Do not do this'	

*This categorization for our consensus document should not be considered as being directly similar to that used for official society guideline recommendations which apply a classification (I–III) and level of evidence (A, B, and C) to recommendations.

green heart indicates recommended strategies, a yellow heart something that can be considered, and a red heart something to be avoided.

This categorization for our consensus document should not be considered as being directly similar to that used for official society guideline recommendations which apply classification (I–III) and level of evidence (A, B, and C) to recommendations.

Data sources

A selection of common and important data sources follow and Table 2 highlights their main strengths and weaknesses. It should be noted that the categories are not completely independent with considerable overlap in some regions.

Registries for regulatory sponsored studies

Registries play an important role in the evaluation of safety and effectiveness of medical devices and pharmaceutical agents. In the case of pharmaceuticals, these registries are also referred to as Phase IV observational studies, which gather information on drug safety and effectiveness after regulatory approval. Regulatory agencies such as the U.S. Food and Drug Administration (FDA) may request a registry as a condition of approval for a device approved under a premarket approval order. Post-approval registries help assess several aspects of therapeutic interventions, including safety, effectiveness, reliability

Table 2 Strengths and weaknesses of common data sources

Strengths	Weaknesses
Regulatory sponsored studies Arrives early after marketing Targeted data collection	Patient selection may not be representative
Learned society academic studies Targeted data collection Usually wide geographical representation	Patient selection need not be representative Quality of outcome registration can vary
Nationwide or regional registries Large scale Less bias in patient selection Low cost	Data quality may be limited given use of clinical documentation International generalizability uncertain
Claims data Complete selection of data within an administrative unit Low cost	Many clinically important data (both independent and outcome variables) may not be available Quality of data may be limited
Investigator-initiated and industry-sponsored studies Multiple centres Careful monitoring of data collected Targeted data collection	Reimbursement for participation can influence patients who consent to intervention Centre selection can result in unrepresentative patients Questions may be designed to ensure a higher probability of a favourable outcome
Hospital cohorts Uniform patient selection Similar expertise to all patients	Patient selection not representative Data quality may not be high Expertise of selected centres may not be generalized

in clinical practice or 'real world' settings, and long-term outcomes. The European Medicines Agency (EMA) launched an initiative for patient registries in 2015 to support more systematic approach to their conduct and use in estimating benefit-risk assessment for pharmaceutical agents in the European Economic Area. Similarly, the EMA also established a European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) and an associated registry database to synergize registry efforts. The ENCePP has also published a Guide on Methodological Standards in Pharmacoepidemiology. (http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml).

There has also been particular interest in the use of registry data to help monitor post-market performance of medical devices.⁴ The FDA has established the unique identifier (UDI) system to incorporate UDI into electronic health information in order to help track individual devices and facilitate tracking outcomes so as to improve nationwide surveillance of device performance. However, the approach to integrating the UDI into data sources has not been established. The FDA is also promoting the development of national and international device registries in several therapeutic areas and interventions. A relevant programme is the National Cardiovascular Data Registry for Implantable Cardioverter-Defibrillators (NCDR ICD, www.ncdr.com). This registry was developed in conjunction with Centers for Medicare and Medicaid Services (CMS) to serve coverage with evidence development decision for primary prevention defibrillators in CMS beneficiaries. This programme has also been employed by FDA and industry for post-market analysis. The NCDR Left atrial appendage occlusion (LAAO) Registry (www.ncdr.com) was also developed in conjunction with FDA and CMS both to fulfil post-marketing requirements (of FDA) and coverage with evidence development (for CMS).

Registries sponsored by learned societies

The EURObservational Research Programme on Atrial Fibrillation (EORP-AF) was an independent initiative promoted by ESC in order to systematically collect data regarding the management and treatment of AF in ESC member countries. The first registry (EORP-AF Pilot Survey) enrolled 3119 patients in 67 centres from February 2012 to March 2013 and showed that the uptake of oral anticoagulation (mostly vitamin K antagonist therapy) had improved since the Euro Heart Survey performed 10 years before, although antiplatelet therapy (especially aspirin) was still used in one-third of the patients and elderly patients were commonly undertreated with oral anticoagulation.⁵⁻⁷ Follow-up data showed that 1-year mortality and morbidity remained high in AF patients, particularly in patients with heart failure or chronic kidney disease.^{7,8} Additionally, asymptomatic AF was particularly common (around 40% of patients) and associated with elderly age, more comorbidities, an high thromboembolic risks and a higher 1-year mortality as compared with symptomatic patients.⁹ As a consequence of the characteristics of the registry, some centres did not participate in long-term follow-up, so only 2119 (68%) patients were included in the 3-year follow-up analysis.¹⁰

The second EORP registry was the EORP-AF Long-Term General Registry, a prospective, observational, large-scale multicentre registry of ESC, that enrolled more than 11 000 AF patients in 250 centres from 27 participating ESC countries from October 2013 to September 2016.¹¹ This registry showed that around 85% of AF

patients are currently treated with oral anticoagulants, with an increase as compared to the past mostly due to the progressive uptake of NOACs.^{11,12} Overall, the registries promoted by ESC over a decade allowed to document significant changes in AF epidemiology in Europe, with an increased complexity of AF patients due to comorbidities, with an impact on both morbidity and mortality.¹²

The American College of Cardiology's PINNACLE Registry is an outpatient, longitudinal clinical quality programme that captures data from ambulatory electronic health records among cardiovascular practice across the USA, and some practices from other countries (e.g. Brazil, India). One of the primary patient cohorts is atrial fibrillation. There have been a number of publications on AF patients from PINNACLE. Recent examples include sex differences in the use of oral anticoagulants, showing that women were less likely to receive anticoagulant therapy at all levels of CHA₂DS₂-VASc score¹³; predictors of oral anticoagulant non-prescription in patients with atrial fibrillation and elevated stroke risk, highlighting the prevalence of anti-platelet use¹⁴; and influence of direct oral anticoagulants on rates of oral anticoagulation for atrial fibrillation, demonstrating that the growing use of direct oral anticoagulants is associated with higher overall oral anticoagulation rates in the USA, although significant practice variation still exists.¹⁵ There have also been nascent efforts to collaborate among global professional society AF registries, with initial participants from the USA, Europe, China, Brazil, South Korea, Taiwan, Singapore, Japan, and the Balkan countries, in order to advance global research insights on AF care and outcomes.¹⁶

The First Brazilian Cardiovascular Registry of Atrial Fibrillation (the RECALL study) will assess demographic characteristics and evidence-based practice of a representative sample of patients with AF in Brazil. Results are expected in 2020.¹⁷

Nationwide cohorts

Large population-based studies can inform on the incidence, prevalence, natural history, treatment, correlates, outcomes, and patterns of healthcare utilization. A special type of large population study encompasses the population of an entire nation. Advantages include very large sample size and lack of selection and participation bias. These advantages are enhanced further when the databases are rich in clinical, personal, and risk factor information and when different pieces of information are linked to permit joint analysis. Once the process for data access is established, vast amounts of information can be obtained at minimal cost, especially when additional collection and update of information is carried out routinely for purposes inherent in medical care and/or insurance coverage and reimbursement. Nationwide cohorts differ from 'Claims data' described below by covering all citizens in an entire region as opposed to an insurance provider where the sample to be examined is defined very differently than a region.

Large nationwide registries are further valuable for examining temporal changes over prolonged time.^{18,19} A recent example is analysis of recurrence of AF following ablation in the Danish register.²⁰ For example, Denmark, Taiwan, Sweden, and Korea have well-established and validated nationwide health insurance (NHI) databases, other national dataset resources, and the capacity for cross-linking some of these databases and/or resources for aetiological information, outcomes, and other data. [Supplementary material online, Table S1](#) shows some main features of the national databases

of these countries.^{21–30} Currently, the Nationwide Research Database includes data files containing information on personal characteristics (sex, date of birth, place of residence, details of insurance, and employment); family relationships; details of clinical information, including date, expenditures, and diagnosis related to both inpatient and outpatient procedures; prescription details; examinations; and operations. While these registries differ in length of retrospective period and specific health data information, their primary strengths include lack of use of selection criteria for enrolment and minimal loss to follow-up. Their weakness is generally lack of obviously important factors such as smoking habits, body weight, etc. except for Korea. Korea database contains lifestyle and habits (body weight, height, smoking, alcohol, and exercise), and basic laboratory data including creatinine, and lipid parameters, etc.³¹

By law, all residents of these countries have a unique personal identification number that is used also for tax returns, bank accounts, and all transactions. Thus, NHI Research Database data are linkable to multiple national databases maintained by other departments, including drug prescriptions, registries of births, deaths, households, immunizations, cancer, reportable infectious diseases, and environmental exposures. In addition, the data in the biobank will be linked with Nationwide Research Database data.

While these sources are highly useful it is also important to point out that access is restricted. Each country has legal restraints to who may access the data. While understandable that the world cannot freely access health information on individuals from a whole population it is important to recognize that anyone wishing to challenge a result from these sources can only do so in collaboration with researchers with proper access authorization.

Claims data

Healthcare systems with access to administrative dataset based on claims data provide an opportunity for observational studies. Examples include insurance data in the USA, such as CMS, which is the payer for services for older persons and the disabled. Claims analyses are limited by appropriateness of coding (usually based on ICD-9 or ICD-10 codes) and whether particular individuals maintain enrolment with the same insurer. Studies that merge multiple claims datasets may identify patients that have been included in >1 insurance datasets. Another important limitation is that patients may not be available for follow-up if they change insurance provider. As for nationwide registers, the level of detail is limited to the information collected, and important and granular clinical data are often missing.

The data have been the basis of recent large comparative effectiveness studies on various NOACs vs. warfarin, or against each other using claims data from the USA. Examples include papers that have investigated NOACs vs. warfarin, and for NOAC vs. NOACs from independent academic groups.³² Claims data have also been used by industry-sponsored studies, for example, those by Lip *et al.*³³

Registries from industry-sponsored cohort studies

Industry sponsorship has led to drug-based registries (e.g. XANTUS, XALIA) and disease-based registries (GARFIELD-AF, GLORIA-AF, PREFER in AF, ORBIT-AF, etc.). There are also several examples of government-funded observational multicentre prospective cohort

studies (PROSE-ICD, PREDETERMINE, Long QT registry, etc.). As these are sponsored efforts, the investigator is often reimbursed for including patients into a particular registry or study, so some element of channelling bias is possible. Nonetheless, by design, there would be including selected patients in (also selected) enrolling centres, but has the positive aspect of careful protocol-based follow-up. In addition to these centre patient-based studies, there are a variety of population-based studies that have been utilized to study arrhythmic endpoints (FHS, ARIC, CHS, MESA, WHS, NHS, and REGARDS).

Hospital cohorts (vs. community)

Hospital cohorts are referred to prospective, or retrospective, observational cohort studies of patients with or at risk for arrhythmia or cardiac conditions and usually receiving a specific treatment or intervention (anticoagulants, ablation, devices, surgery, etc.). They may be local cohorts or wider scale regional or national cohorts covering a global healthcare system. Nationwide hospital cohorts can provide real-world evidence of clinical practice, patient outcomes, safety, comparative effectiveness, and cost-effectiveness of interventions. A systematic robust research design, with accurate measurement of appropriate outcomes and control variables, is needed for protecting the quality of data.

Both hospital and community-based cohorts can be used to evaluate the outcomes of patients exposed to a particular programme or management strategy and are useful for understanding the real-world safety and effectiveness of specific treatments and may provide the analysis of the relative effectiveness of a given treatment among alternative patients' subgroups. Compared to hospital cohorts, the communities' cohorts can provide the advantage of longitudinal data collection on considerable number of unselected patients. The key endpoints, such as mortality information, could be attained from the hospital cohort, which are variably missing in administrative claims databases. In contrast, nationwide administrative databases may identify outcomes recorded on different healthcare facilities on a larger scale and may reduce channelling bias (see below).

Hospital cohorts have important limitations. Hospital uptake may be highly selective resulting in patients for study being of higher or lower risk than the average patient. Such weaknesses may also vary over time as treatments change from in-hospital to outpatient treatment.

Bias and confounding

Bias

All studies including randomized studies are potentially subject to processes that may cause a study to report results that may not be generalized or may even be incorrect. These processes are referred to as bias and nearly all bias is related to the selection of the study population (selection bias) or recording of data from a study (information bias). Sacket lists 35 types of bias³⁴ and the list is far from complete. Table 3 is a selected list of either common or commonly overlooked sources of bias.

In addition to bias that can at least be listed as limitations, there are other sources. Data dredging bias is when multiple analyses are performed on a dataset and only the apparently interesting ones are reported. It is related to publication bias, where journals are more likely to accept potentially interesting positive findings, but once an interesting finding has been published the absence of the same finding

Table 3 Selected sources of bias

Bias	Description
Selection bias	Subjects chosen for the study are not representative of the population of interest
Prevalence incidence (Neyman) bias	A late look at those with a disease or condition will miss early problems and those that have died
Admission rate (Berkson) bias	A hospital-based study of the relation between a disease and some exposure will be biased if patients with the disease are more or less admitted to hospital depending on the exposure of interest
Immortal lifetime bias	When future events are included as baseline data those that have the future event will be immortal until the time when the future data were recorded
Unmasking (detection signal) bias	An innocent exposure may become associated with disease if it triggers search for a disease
Volunteer bias	Individuals volunteering for studies or seeking early help for symptoms may be more healthy than non-volunteers or latecomers
Response bias	People who agree to take part in a study have different characteristics from those that do not, and this distorts the results when making conclusions about the whole population
Withdrawal bias	If patients that discontinue a study differ importantly from those that remain in a study the final result may be severely distorted, in particular when only measurements at the end of the study, such as rhythm control can enter the analyses
Channelling bias	The propensity of 'sicker' or selected patients to be prescribed disproportionately the newer and perceived to be more potent medications differentially
Confounding by indication, nearly identical to channelling bias	When studying an intervention such as a pharmaceutical drug it may be impossible to distinguish between the risk of the intervention and the risk of the condition that triggered the intervention
Protopathic bias (reverse causation)	The exposure changes as a result of early disease manifestations. If patients change lifestyle because of early disease signs a wrong direction between lifestyle and disease may be observed
Information bias	
Recall bias	Information that relies on patient memory may be influenced by their condition. If a relation between a disease and a symptom is available to the patient that may help the patient remember a condition
Insensitive measure bias	If the measurement used in a study does not detect what it is supposed to detect and underestimation of that measurement will be the result
Regression dilution bias	If a measurement is inaccurate the relation between the measurement and outcome is weakened. For comparison of continuous variables, the slope will be reduced
Follow-up bias	If follow-up depends on the presence of a condition this can create a false relation between a condition and a disease, the direction depending on whether the condition improves or worsens follow-up
Assessment bias	The assessment and thus collected data on a subject is influence by other factors
Interviewer bias	If an interviewer is aware of the subject's health status, this may influence the questions asked, or how they are asked, which consequently affects the response

may become interesting enough for publication. Cognitive dissonance bias is when strong beliefs prevail in spite of evidence.

So, what can be done about bias? The always important limitations of observational studies are that unknown or unaccounted bias can never be completely excluded. There is no mathematical technique to adjust for bias that is potentially present but not known. On occasion subgroup analyses and other sensitivity analyses may cast light on the problems in a study.

In many cases bias is complex. One example is comparison of treatments and allowing both prevalent and new users in an analysis. This introduces several sources of bias. There is a selection bias towards patients that tolerate a certain therapy and information bias that therapy can change the covariates. A new user design is preferable for examination of the importance of any treatment.³⁵

Confounding

A confounder is classically defined as a factor which influences both the exposure and the outcome. If for example, a study of implantable defibrillators for heart failure is randomized, then we would expect

all characteristics of the patients to be equally distributed in the two groups. Factors such as age and sex would be expected to be (nearly) identical in the two groups. And also factors of importance that we do not know (unknown confounders/residual confounders) would be expected to be similar in the two groups. If, on the other hand, the study was observational, then we would expect age and sex to be differently distributed between the two groups. Age and sex would also be expected to be important for survival. In this case, age and sex are examples of the classical definition of a confounder: they are unevenly distributed between the treatment groups and they have importance for the outcome.

Classical confounders such as age and sex are accounted for by including them as covariates in a multivariable model. The distinction between confounders and model covariates can easily become blurred. Usually, we have to select a reasonable number of known factors as potential confounders and use them as covariates in analysis. Directed acyclic graphs ([Supplementary material online](#)) are often a helpful instrument. For example, socioeconomic status of patients could also influence survival and in an observational study

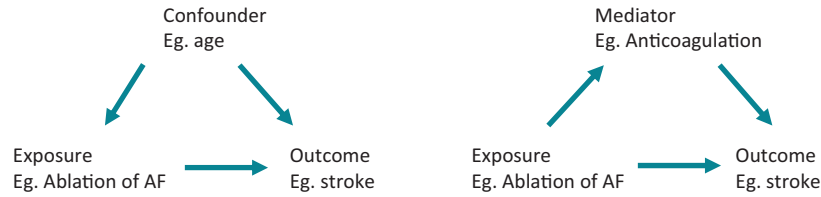


Figure 1 Directed acyclic graphs of a confounder and a mediator. AF, atrial fibrillation.

socioeconomic status could also influence whether a patient received a defibrillator. If we do not have a recording of socioeconomic status it would be a classical example of an unknown confounder. Ultimately, all observational analyses are potentially subject to bias from unknown confounders.

If we further have a recording of myocardial infarction after implantation, such a variable should not be used in analysis of the importance of the defibrillator. First, the infarction comes after study start. A patient obviously cannot die before the infarction, and therefore, an immortal lifetime bias is introduced in a simple analysis. Further, the infarction lies on the pathological pathway between having a defibrillator and the outcome of mortality. It is an intermediate and intermediates should not be used as confounder. Because of its position on the pathway between defibrillator and death, it might distort the result if by some mechanism there was an association between getting a defibrillator and the risk of a myocardial infarction. For a more technical approach to confounding, we refer to previous literature.^{36,37}

Mediation

A mediator or intermediate variable is a variable/factor which lies on the pathological path between the exposure of interest and the outcome. Figure 1 shows the major difference between a mediator and a confounder. Appropriate analysis of mediators is complex and there is further explanation in the [Supplementary material](#) online, *Appendix*. Mediators should not be treated as confounders.

Causal inference

Causal inference is a framework to derive average treatment effects from observational studies with the ultimate aim (or hope) of demonstrating a causal interpretation. If the above study of defibrillators to patients with heart failure was randomized, and we after a year found that the mortality with a defibrillator was 4% and 7% without a defibrillator. We could then calculate the *average treatment effect* at 1 year of 3%. Assuming that the trial was also statistically significant that average treatment effect would be a very important message and easily used to calculate the number of patients to treat to save a life (over 1 year).

On the other hand, if our study was observational, we might also have a difference in mortality of 3% after 1 year. But we would have age, sex, and other factors being different in the two groups, so we

could not expect the 3% to hold for the average patient even if we have no unknown confounders. We could present a multivariable model with hazard ratios or odds ratios, but the average treatment effect from the randomized trial and the number needed to treat would not be available.

Causal inference is a framework to derive the average treatment effect of an observational study providing that we have perfect adjustment for *all* confounders. From a clinical perspective, two methods from causal inference are useful and used: *Propensity adjustment* and the *G-formula*. The reader interested in further detail including formal assumptions is referred to an excellent book on the subject: 'Causal inference'.³⁸

In the case of *propensity score matching*, using regression analysis, we would calculate the 'propensity' for getting a defibrillator for the entire cohort, including those with and without a defibrillator. This is simply the probability of getting a defibrillator given the covariates. We would then match patients with and without a defibrillator as having very similar probability of getting one. We would discard patients from the analysis when they cannot be reasonably matched. When the technique is successful, we have a moderately smaller sample than we started with and a demographic table that shows similar covariate distribution in both groups. We can then use the same instruments as we used in the randomized study to obtain *average treatment effect* (actually average treatment effect of the treated) and number needed to treat. The *pitfalls* of this method arrive when the covariates actually do not predict treatment and the demographic table after matching does not show a good balance.

Causal inference provides average treatment effects as do randomized studies, but observational studies are not randomized, and therefore, the presence of unknown or unmeasured confounders may drive differences. Only large randomized studies assure control of unmeasured confounders.

A technique related to propensity score matching is *inverse probability weighting*. With this technique, cases are given a weight corresponding their probability of receiving the treatment of interest. This technique can also provide average treatment effect. It has the advantage that all patients are included in the analysis.³⁹

While propensity matching is commonly used it has the important disadvantage that not all patients can be matched and commonly not all covariates are evenly distributed after matching. Another technique that has become available is to simulate a randomized trial where first *all* the patients in the study receive a defibrillator and afterwards all patients do not get a defibrillator. This technique is called the *G-formula* and it relies on using statistical models to predict the

Table 4 Common epidemiological modelling methods

Model	Description	Critical assumptions
Cox proportional hazard	Models risk as hazard ratio, there is a single non-parametric time scale	Proportional hazard assumption—the ratio between hazards needs to be constant
Poisson regression	Time is split into interval as dependent of up to many time scales and timing of covariates	The rate of events needs to be constant in intervals
Logistic regression	Examines only the outcome as usually a bivariate outcome	Can be used in outcome studies when there is no censoring
G-modelling	Causal inference—one of the above models is used to predict outcome at a time point for the WHOLE study population	Simulates a randomized experiment where the whole study population is subjected to all treatments—assumes no residual confounding
Matching on covariates prior to modelling	Reduces modelling assumptions by perfect adjustment for the matching covariates. The sample size may be reduced	Requires that the selected covariates define necessary confounding and lack of important unknown confounders
Propensity stratified models	Uses covariates to calculate the probability of receiving one of two treatments and then compares outcome in strata of that probability	Assumes that the difference in treatment is perfectly explained by the probability of receiving treatment
Propensity-matched models	The propensity is calculated as above and then cases with same or very similar probability in two groups are matched	Same as above, depending on the matching the sample size may be reduced

outcome of every patient first with a defibrillator and then without a defibrillator. Using this simulated study, we can calculate *average treatment effect* and number needed to treat using suitable techniques.³⁸ In propensity score matching of the defibrillator study it was a requirement that the covariates predict whether a patient gets a defibrillator. The G-model does not have this requirement, but the requirement that the covariates predict the outcome accurately and that there are no unknown confounders.

The G-formula and propensity-based techniques are not competing techniques, but each has advantages and disadvantages—and both allow calculation of average treatment effects and numbers needed to treat.

Statistical modelling

Addressing again an observational study of defibrillators to patients with heart failure, we would expect to find that age, sex, and other variables would differ among patients with and without a defibrillator. The most basic technique for handling this is stratification—to study independently young vs. old and men vs. women, etc. This is useful if there are few variables with few values which is rarely the case. Another technique is to match patients with and without defibrillators and having the same age, sex, etc. This is a very efficient technique but usually fails because it is not possible to find a match for many patients. Instead of matching on each variable, we could turn to propensity score matching above which may or may not solve our matching problem.

The alternative to matching and stratification is a statistical model and Table 4 lists commonly used models. Such models output

parameter estimates which after transformation provide odds ratios, hazard ratios, or rate ratios. If these measures are statistically significant there is an association between a factor of interest and the outcome of interest. This may be entirely useful for a study of whether a factor has some importance for an outcome, but it is important to realize that this importance cannot be interpreted as prediction. It is therefore important to determine whether the object of a study is to explain or to predict.⁴⁰ Some uncertainty arises from the fact that 'risk' and 'prediction' do not have universally defined mathematical equivalents. For the current account, *prediction* is defined as the absolute risk at a defined time horizon. There is a recent example from the hypertension field.⁴¹ This study used hazard ratios to argue for a value of ambulatory blood pressure, but the aim was to examine whether ambulatory blood pressure improved prediction of cardiovascular outcomes. When encouraged to actually calculate a change in prediction the actual improvement in predictive value was very small.^{41,42} For a study of this nature it would be natural to focus on predictive value rather than on hazard ratios.⁴³ There is plenty of literature to show that even very high or low hazard ratios may have little relation to prediction.^{44–48} In general, whenever the importance of a new treatment or a new biomarker is involved it should be considered whether prediction is the more important estimate to calculate.

C-index/area under a receiver operator curve

Let us assume that we want to examine whether late potentials add to prediction of cardiovascular mortality in patients with heart failure. A simple approach would be to present the hazard ratio of some cut-off of late potentials. If this was significant, we could assume late

potentials to have some importance. But as described above in the section on hazard ratio and below with competing risk, we would not have assurance that we can predict cardiovascular mortality at 5 years. The right method to show the benefit of a ‘new’ biomarker such as the suggested late potentials demonstrate that a properly selected C-index or area under a receiver operator curve is significantly changed by a new biomarker.^{44,46} This is a field in development with several pitfalls. Thus, the commonly used methods of integrated discrimination improvement (IDI) and net reclassification index (NRI)⁴⁹ are not valid. Addition of random data to datasets can improve the parameters. The C-index from a Cox model should also not be used to indicate discriminative improvement at specific times.⁵⁰

The bottom line for selection of statistical models is to ensure such a discussion between statisticians and clinicians that the statistical methods used match the clinical question. If the aim is to estimate the survival benefit of a defibrillator in heart failure after 5 years then a model that address prediction should be used. If it is sufficient to know that the defibrillator does ‘something’, then models that provide hazard ratio, rate ratio, or odds ratio may suffice.

Competing risk

Let us assume in the study of defibrillators for heart failure that we were not so much interested in all-cause mortality but rather in cardiovascular mortality. This would not be unreasonable since defibrillators can only influence cardiovascular mortality. This has important consequences for the analysis. The competing risk of death from other causes than cardiovascular mortality cannot be ignored and the cumulative cardiovascular mortality presentation needs to take into account the competing risk with proper technique.⁵¹

Competing risk has for technical reasons no influence on the calculation of hazard ratios, but the interpretation of hazard ratio becomes complex. In fact, there is no certainty that a significant hazard ratio influences long-term prediction such as 5-year cardiovascular mortality and dedicated analysis of prediction is necessary if this is the goal.

Instrumental variable analysis

A good instrument is a variable that affects an outcome and is not affected by confounders. The only common example in clinical medicine is ‘Mendelian randomization’. With this technique, genes that influence a factor of interest is used instead of directly addressing the factor. Since genes have been there prior to establishing the influence of important confounders that could be age and smoking the confounding by these can be avoided. More detail is provided in the [Supplementary material](#) online, *Appendix*. It is important to appreciate the limitations and a good reference is Federspiel *et al.*⁵²

Missing data

Missing data are common in observational studies and most statistical procedures exclude individuals with missing data. If in the study of defibrillators for heart failure and important variable such as age is missing for some patients it could bias the interpretation of the study if these patients are simply removed from the analysis. There are a

number of useful techniques to include as much information as possible from cases with missing data and these are described further in the [Supplementary material](#) online, *Appendix*.

Common problems

Causality vs. association

Observational studies will by their nature always include a risk of bias from unknown or unobserved confounders. Causal language is common and a very common task for reviewers is to request the removal of causal language from observational manuscripts. It can be argued that in stating the objective of a study a causal language should be used.⁵³

Conditioning on the future

Conditioning on the future is when information is obtained sometime in the future compared to baseline is included as baseline information. Patients that pick up a prescription cannot die before that day, while patients dying prior to reaching the pharmacy never pick up a prescription. Using the prescription information at baseline will bias survival towards those that pick up a prescription—the immortal lifetime bias.⁵⁴ It is a very similar problem if patients are excluded from a study because of events after baseline—this will in a very similar manner bias survival towards those that do not have the factor that caused exclusion. Friberg *et al.*^{55,56} studied stroke in atrial fibrillation not treated with anticoagulation. By excluding patients who received anticoagulation during the study a bias was introduced. This particular bias was examined in a different study⁵⁷ that demonstrated a bias towards lower stroke rate with low CHA₂DS₂-VASc by excluding after baseline.

Meta-analysis of observational studies

Meta-analyses of RCTs assume that each individual study provides an unbiased estimate of the effect and any variability between study results is attributed to random variation.^{58,59} The overall effect will provide an unbiased estimate, as long as the studies are representative and wisely combined.^{58,59} While RCTs, if properly designed, are expected to have a high internal validity, they traditionally have the limitations of smaller sample sizes, very selected populations, shorter follow-up time, ethical constraints, and high cost.^{60,61} Incorporating non-randomized trials into meta-analyses can overcome some of these limitations by improving generalizability (more diverse populations), allowing larger sample sizes, allowing exploring aetiological hypothesis (unethical to deliberately expose patients to harmful risk factors in an RCT), and evaluating less common adverse effects.^{60–62}


Observational studies, however, have a higher risk of bias and confounding and, as a consequence, the association estimates may differ from the truth beyond the effect of chance.^{63,64} The individual studies may measure and control for known confounding factors during the analysis. However, even if this is case, bias and residual confounding (i.e. when the confounding factor cannot be measured with sufficient precision^{65,66}) remain a relevant threat to validity in observational research.⁶⁷ As a consequence, using non-randomized studies in meta-

analysis could (more often than not) perpetuate the biases that are unknown, unmeasured or uncontrolled in these observational studies, and threaten the validity of the entire meta-analysis.^{64,67,68} Furthermore, reporting in observational studies is frequently not sufficiently detailed to judge their limitations,^{67,69–71} they show significant heterogeneity,^{72–74} and deficiencies in methodology.^{68,75,76} Network meta-analyses (i.e. meta-analyses that compare simultaneously multiple treatment options) incorporating non-randomized trials face similar challenges.⁷⁷

For these reasons, some authors recommend abandoning meta-analyses of observational data.^{64,78,79} Yet, when evaluating effect sizes derived from meta-analyses of RCTs and non-randomized studies, discrepancies have shown to be small in high-quality observational studies with little heterogeneity.^{60,80–83} Still, discrepancies beyond chance do happen, and it is, therefore, essential to assess the differences between studies.^{61,64} In our—and other authors'—view, gross statistical combination of data alone should be avoided; rather, a thorough analysis of heterogeneity sources and possible bias should be done^{61,73,84,85}; this will probably provide better understanding than an overall effect measure, which can potentially be misleading.⁷³

In 1999, the *Quality of Reporting of Meta-analyses* (QUOROM) statement was issued 'to address standards for improving the quality of reporting of meta-analyses of RCTs'.⁸⁶ A similar checklist was published in 2000 for reporting *Meta-analyses Of Observational Studies in Epidemiology* (MOOSE).⁷³ However, in the face of persistent poor reporting,^{69,70,87–94} these statements were later on updated in the form of the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) statements.^{95–101} Many peer-reviewed journals now require that these guidelines are followed when submitting a systematic review or meta-analyses, as the endorsement of these statements improves both reporting and methodological quality^{102,103}; however, there is still room for improvement.^{104–107} For editors, reviewers, and readers, a measurement tool to assess the methodological quality of systematic reviews (AMSTAR) has also been published and validated.^{108–110}

Consensus statements on observational studies

		Refs
<p>When reporting the results of an observational study/meta-analysis, the STROBE/PRISMA statement checklists should be used:</p> <p>STROBE—www.strobe-statement.org</p> <p>PRISMA—www.prisma-statement.org</p>		

Continued

Continued

		Refs
<p>Prior to analysis an analysis, plan should be agreed upon and formally recorded</p>		<p>www.strobe-statement.org</p>
<p>The process of data collection should be clearly presented so that the strengths and limitations are clear to the reader</p>		
<p>If legally possible data should be available for scrutiny by other researchers</p>		
<p>Studies should have clear objective and use statistical methods that match the objectives</p>		<p>38</p>
<p>The reporting of findings should be complete and the strengths and limitations clearly described</p>		
<p>Sources of bias should be identified and presented to the reader</p>		

Conclusion

Observational studies should in general use transparent and valid methodology and use concise reporting. There are available guidelines for epidemiological studies and the most recent is from the International Society of Pharmacoepidemiology.¹¹¹ The guideline from the International Society of Pharmacoepidemiology also cites a number of other guidelines. None of the recommendations are in discordance with the current consensus statement. There does not appear to be widely accepted international guidelines for 'good epidemiological practice'.¹¹² Finally, an important intermediate step is to ensure that biostatisticians and clinical practitioners both have sufficient insight into the language and methods of each other to ensure that valid studies are conducted and the many pitfalls avoided.

Acknowledgements

The authors thank ESC Scientific Document Group: Dr. Nikolaos Dages, Dr. Serge Boveda, Dr. Kevin Vernooij, Prof. Zbigniew Kalarus, Prof. Gulmira Kudaiberdieva, Dr. Georges H Mairesse, Prof. Valentina Kutiyifa, Prof. Thomas Deneke, Prof. Jesper Hastrup Svendsen, Dr. Vassil B Traykov, Prof. Arthur Wilde, Prof. Frank R. Heinzel.

Supplementary material

Supplementary material is available at *Europace* online.

Conflict of interest: Christian Torp-Pedersen reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Bayer. Gregory Lip reports

Consultancy for Bayer/Janssen, BMS/Pfizer, Medtronic, Boehringer Ingelheim, Novartis, Verseon and Daiichi-Sankyo; Speaker for Bayer, BMS/Pfizer, Medtronic, Boehringer Ingelheim, and Daiichi-Sankyo. No fees are directly received personally. Christine Albert reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Myocardia and Sanofi Aventis; Research funding from Roche Diagnostics : Biomarkers and St Jude Medical. Elena Arbelo reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Biosense Webster. Alvaro Avezum reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Bayer and Boehringer Ingelheim. Giuseppe Boriani reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Boston Scientific, Medtronic and Biotronik. John Camm reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from BMS, Sorin, Wiley Blackwell, Boehringer Ingelheim, Oxford University Press, GSK, InCardia, Milestone, Menarini, Bayer, Medtronic; Research funding from Radius, Thrombosis Research Institute, BMS/Pfizer, Daiichi Sankyo, Servier, Armethon, Richmond Pharmacology, Cardiac Insight.

Laurent Fauchier reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Boehringer Ingelheim, Daiichi Sankyo, Berlin Chemie AG, Bayer, Medtronic. Young-Hoon Kim reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Medtronic, Abbott, Daiichi Sankyo, Pfizer, Bayer. Frederick Masoudi reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from American Board of Internal Medicine and the Massachusetts Medical Society. Peter B Nielsen reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Pfizer and Bayer; Research funding from BMS. Jonathan Piccini reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Sanofi, Allergan, Phillips, Medtronic; Research funding from Abbott, Gilead, Janssen, ARCA biopharma, Boston Scientific. Tatjana Potpara reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Bayer and Pfizer. Flemming Skjøth reports Speaker fees, Honoraria, Consultancy, Advisory Board fees, Investigator, Committee Member, etc. from Bayer. Other authors: none declared.

References

1. Faranoff AC, Califf RM, Windecker S, Smith SC Jr, Lopes RD *et al.* Levels of evidence supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. *JAMA* 2019;**321**: 1069-80.
2. Frieden TR. Evidence for health decision making—beyond randomized, controlled trials. *N Engl J Med* 2017;**377**:465-75.
3. Shikata S, Nakayama T, Noguchi Y, Taji Y, Yamagishi H. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. *Ann Surg* 2006;**244**:668-76.
4. Blake K. Postmarket surveillance of medical devices: current capabilities and future opportunities. *J Interv Card Electrophysiol* 2013;**36**:119-27.
5. Lip GYH, Laroche C, Dan G-A, Santini M, Kalarus Z, Rasmussen LH *et al.* A prospective survey in European Society of Cardiology member countries of atrial fibrillation management: baseline results of EURObservational Research

- Programme Atrial Fibrillation (EORP-AF) Pilot General Registry. *Europace* 2014;**16**:308-19.
6. Lip GYH, Laroche C, Dan G-A, Santini M, Kalarus Z, Rasmussen LH *et al.* 'Real-world' antithrombotic treatment in atrial fibrillation: the EORP-AF pilot survey. *Am J Med* 2014;**127**:519-29.e1.
7. Lip GYH, Laroche C, Ioachim PM, Rasmussen LH, Vitali-Serdoz L, Petrescu L *et al.* Prognosis and treatment of atrial fibrillation patients by European cardiologists: one year follow-up of the EURObservational Research Programme-Atrial Fibrillation General Registry Pilot Phase (EORP-AF Pilot registry). *Eur Heart J* 2014;**35**:3365-76.
8. Boriani G, Laroche C, Diemberger I, Popescu MI, Rasmussen LH, Petrescu L *et al.* Glomerular filtration rate in patients with atrial fibrillation and 1-year outcomes. *Sci Rep* 2016;**6**:30271.
9. Boriani G, Laroche C, Diemberger I, Fantecchi E, Popescu MI, Rasmussen LH *et al.* Asymptomatic atrial fibrillation: clinical correlates, management, and outcomes in the EORP-AF Pilot General Registry. *Am J Med* 2015;**128**:509-18.e2.
10. Boriani G, Proietti M, Laroche C, Diemberger I, Popescu MI, Riahi S *et al.* Changes to oral anticoagulant therapy and risk of death over a 3-year follow-up of a contemporary cohort of European patients with atrial fibrillation final report of the EURObservational Research Programme on Atrial Fibrillation (EORP-AF) pilot general registry. *Int J Cardiol* 2018;**271**:68-74.
11. Boriani G, Proietti M, Laroche C, Fauchier L, Marin F, Nabauer M *et al.* Contemporary stroke prevention strategies in 11 096 European patients with atrial fibrillation: a report from the EURObservational Research Programme on Atrial Fibrillation (EORP-AF) Long-Term General Registry. *Europace* 2018;**20**: 747-57.
12. Proietti M, Laroche C, Nieuwlaet R, Crijns HJGM, Maggioni AP, Lane DA *et al.* Increased burden of comorbidities and risk of cardiovascular death in atrial fibrillation patients in Europe over ten years: a comparison between EORP-AF pilot and EHS-AF registries. *Eur J Intern Med* 2018;**55**:28-34.
13. Thompson LE, Maddox TM, Lei L, Grunwald GK, Bradley SM, Peterson PN *et al.* Sex differences in the use of oral anticoagulants for atrial fibrillation: a report from the National Cardiovascular Data Registry (NCDR((R))) PINNACLE Registry. *J Am Heart Assoc* 2017;**6**:1-10.
14. Lubitz SA, Khurshid S, Weng L-C, Doros G, Keach JW, Gao Q *et al.* Predictors of oral anticoagulant non-prescription in patients with atrial fibrillation and elevated stroke risk. *Am Heart J* 2018;**200**:24-31.
15. Marzec LN, Wang J, Shah ND, Chan PS, Ting HH, Gosch KL *et al.* Influence of direct oral anticoagulants on rates of oral anticoagulation for atrial fibrillation. *J Am Coll Cardiol* 2017;**69**:2475-84.
16. Hsu JC, Akao M, Abe M, Anderson KL, Avezum A, Gluskamp N *et al.* International Collaborative Partnership for the Study of Atrial Fibrillation (INTERAF): rationale, design, and initial descriptives. *J Am Heart Assoc* 2016;**5**: 1-14.
17. Lopes RD, de Paola AAV, Lorga Filho AM, Consolim-Colombo FM, Andrade J, Piva e Mattos LA *et al.* Rationale and design of the First Brazilian Cardiovascular Registry of Atrial Fibrillation: the RECALL study. *Am Heart J* 2016;**176**:10-16.
18. Kim D, Yang P-S, Jang E, Yu HT, Kim T-H, Uhm J-S *et al.* 10-year nationwide trends of the incidence, prevalence, and adverse outcomes of non-valvular atrial fibrillation nationwide health insurance data covering the entire Korean population. *Am Heart J* 2018;**202**:20-6.
19. Kim D, Yang PS, Jang E, Yu HT, Kim TH, Uhm JS *et al.* Increasing trends in hospital care burden of atrial fibrillation in Korea, 2006 through 2015. *Heart* 2018;**104**:2010-17.
20. Pallisgaard JL, Gislason GH, Hansen J, Johannessen A, Torp-Pedersen C, Rasmussen PV *et al.* Temporal trends in atrial fibrillation recurrence rates after ablation between 2005 and 2014: a nationwide Danish cohort study. *Eur Heart J* 2018;**39**:442-9.
21. Nielsen PB, Larsen TB, Skjøth F, Lip GYH. Outcomes associated with resuming warfarin treatment after hemorrhagic stroke or traumatic intracranial hemorrhage in patients with atrial fibrillation. *JAMA Intern Med* 2017;**177**:563-70.
22. Nielsen PB, Skjøth F, Søgaard M, Kjældgaard JN, Lip GY, Larsen TB *et al.* Effectiveness and safety of reduced dose non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study. *BMJ* 2017;**356**:j510.
23. Kim TH, Yang PS, Uhm JS, Kim JY, Pak HN, Lee MH *et al.* CHA2DS2-VASc Score (Congestive Heart Failure, Hypertension, Age \geq 75 [Doubled], Diabetes Mellitus, Prior Stroke or Transient Ischemic Attack [Doubled], Vascular Disease, Age 65-74, Female) for Stroke in Asian Patients With Atrial Fibrillation: a Korean Nationwide Sample Cohort Study. *Stroke* 2017;**48**: 1524-30.
24. Kim T-H, Yang P-S, Kim D, Yu HT, Uhm J-S, Kim J-Y *et al.* CHA2DS2-VASc score for identifying truly low-risk atrial fibrillation for stroke: a Korean Nationwide Cohort Study. *Stroke* 2017;**48**:2984-90.

25. Savarese G, Sartipy U, Friberg L, Dahlström U, Lund LH. Reasons for and consequences of oral anticoagulant underuse in atrial fibrillation with heart failure. *Heart* 2018;**104**:1093–100.
26. Karayiannides S, Lundman P, Friberg L, Norhammar A. High overall cardiovascular risk and mortality in patients with atrial fibrillation and diabetes: a nationwide report. *Diab Vasc Dis Res* 2018;**15**:31–8.
27. Chao T-F, Lip GYH, Lin Y-J, Chang S-L, Lo L-W, Hu Y-F et al. Major bleeding and intracranial hemorrhage risk prediction in patients with atrial fibrillation: attention to modifiable bleeding risk factors or use of a bleeding risk stratification score? A nationwide cohort study. *Int J Cardiol* 2018;**254**:157–61.
28. Chao T-F, Lip GYH, Liu C-J, Lin Y-J, Chang S-L, Lo L-W et al. Relationship of aging and incident comorbidities to stroke risk in patients with atrial fibrillation. *J Am Coll Cardiol* 2018;**71**:122–32.
29. Chao T-F, Liu C-J, Tuan T-C, Chen T-J, Hsieh M-H, Lip GYH et al. Lifetime risks, projected numbers, and adverse outcomes in Asian patients with atrial fibrillation: a report from the Taiwan Nationwide AF Cohort Study. *Chest* 2018;**153**:453–66.
30. Hsing AW, Ioannidis JP. Nationwide population science: lessons from the Taiwan National Health Insurance Research Database. *JAMA Intern Med* 2015;**175**:1527–9.
31. Lee S-R, Choi E-K, Han K-D, Jung J-H, Oh S, Lip GYH et al. Edoxaban in Asian patients with atrial fibrillation: effectiveness and safety. *J Am Coll Cardiol* 2018;**72**:838–53.
32. Noseworthy PA, Yao X, Shah ND, Gersh BJ. Comparative effectiveness and safety of non-vitamin K antagonist oral anticoagulants versus warfarin in patients with atrial fibrillation and valvular heart disease. *Int J Cardiol* 2016;**209**:181–3.
33. Lip GYH, Keshishian A, Li X, Hamilton M, Masseria C, Gupta K et al. Effectiveness and safety of oral anticoagulants among nonvalvular atrial fibrillation patients. *Stroke* 2018;**49**:2933–44.
34. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979;**32**:51–63.
35. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;**158**:915–20.
36. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia: Wolters Kluwer; 2012.
37. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;**15**:615–25.
38. Hernán MA, Robins JM. *Causal Inference*. Boca Raton: Chapman & Hall/CRC; 2018.
39. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015;**34**:3661–79.
40. Shmueli G. To explain or to predict. *Stat Sci* 2010;**25**:289–310.
41. Banegas JR, Ruilope LM, de la Sierra A, Vinyoles E, Gorostidi M, de la Cruz JJ et al. Relationship between clinic and ambulatory blood-pressure measurements and mortality. *N Engl J Med* 2018;**378**:1509–20.
42. Torp-Pedersen C. Ambulatory blood pressure and mortality. *N Engl J Med* 2018;**379**:1285–6.
43. Mortensen RN, Gerds TA, Jeppesen JL, Torp-Pedersen C. Office blood pressure or ambulatory blood pressure for the prediction of cardiovascular events. *Eur Heart J* 2017;**38**:3296–304.
44. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MSV et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation* 2009;**119**:2408–16.
45. Kattan MW. Judging new markers by their ability to improve predictive accuracy. *J Natl Cancer Inst* 2003;**95**:634–5.
46. Kattan MW. Evaluating a new marker's predictive contribution. *Clin Cancer Res* 2004;**10**:822–4.
47. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004;**159**:882–90.
48. Hernan MA. The hazards of hazard ratios. *Epidemiology* 2010;**21**:13–15.
49. Hilden J, Gerds TA. A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and net reclassification index. *Stat Med* 2014;**33**:3405–14.
50. Blanche P, Kattan MW, Gerds TA. The c-index is not proper for the evaluation of t-year predicted risks. *Biostatistics* 2018;**20**:347–357.
51. Kleinbaum DG, Klein M. Competing risks survival analysis. In Gail M, Samet JM. *Survival Analysis: A Self-Learning Text*. Springer: New York, 2012. pp. 425–95.
52. Federspiel JJ, Anstrom KJ, Xian Y, McCoy LA, Effron MB, Faries DE et al. Comparing inverse probability of treatment weighting and instrumental variable methods for the evaluation of adenosine diphosphate receptor inhibitors after percutaneous coronary intervention. *JAMA Cardiol* 2016;**1**:655–65.
53. Hernan MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health* 2018;**108**:616–9.
54. Lund J, Horváth-Puhó E, Komjáthi Szépligeti S, Sørensen HT, Pedersen L, Ehrenstein V et al. Conditioning on future exposure to define study cohorts can induce bias: the case of low-dose acetylsalicylic acid and risk of major bleeding. *Clin Epidemiol* 2017;**9**:611–26.
55. Friberg L, Skeppholm M, Terent A. Benefit of anticoagulation unlikely in patients with atrial fibrillation and a CHA₂DS₂-VASc score of 1. *J Am Coll Cardiol* 2015;**65**:225–32.
56. Aspberg S, Chang Y, Atterman A, Bottai M, Go AS, Singer DE et al. Comparison of the ATRIA, CHADS₂, and CHA₂DS₂-VASc stroke risk scores in predicting ischaemic stroke in a large Swedish cohort of patients with atrial fibrillation. *Eur Heart J* 2016;**37**:3203–10.
57. Nielsen PB, Larsen TB, Skjøth F, Overvad TF, Lip GY. Stroke and thromboembolic event rates in atrial fibrillation according to different guideline treatment thresholds: a nationwide cohort study. *Sci Rep* 2016;**6**:27410.
58. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res* 2001;**10**:277–303.
59. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In JPT Higgins and S Green (eds). *Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0)*, 2011. http://handbook-5-1.cochrane.org/chapter_9/9_analysing_data_and_undertaking_meta_analyses.htm (16 September 2019, date last accessed).
60. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;**342**:1887–92.
61. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;**286**:821–30.
62. Lipsett M, Campleman S. Occupational exposure to diesel exhaust and lung cancer: a meta-analysis. *Am J Public Health* 1999;**89**:1009–17.
63. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002;**359**:248–52.
64. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;**7**:iii–x, 1–173.
65. Phillips AN, Smith GD. How independent are “independent” effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol* 1991;**44**:1223–31.
66. Smith GD, Phillips AN. Confounding in epidemiological studies: why “independent” effects may not be all they seem. *BMJ* 1992;**305**:757–9.
67. Groenwold RHH, Van Deursen AMM, Hoes AW, Hak E. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. *Ann Epidemiol* 2008;**18**:746–51.
68. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6.
69. Hemels MEH, Vicente C, Sadri H, Masson MJ, Einarson TR. Quality assessment of meta-analyses of RCTs of pharmacotherapy in major depressive disorder. *Curr Med Res Opin* 2004;**20**:477–84.
70. Dixon E, Hameed M, Sutherland F, Cook DJ, Doig C. Evaluating meta-analyses in the general surgical literature: a critical appraisal. *Ann Surg* 2005;**241**:450–9.
71. Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007;**147**:1500–24.
72. Maguire MJ, Hemming K, Hutton JL, Marson AG. Overwhelming heterogeneity in systematic reviews of observational anti-epileptic studies. *Epilepsy Res* 2008;**80**:201–12.
73. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000;**283**:2008–12.
74. Int'Hout J, Ioannidis JPA, Borm GF, Goeman JJ. Small studies are more heterogeneous than large ones: a meta-meta-analysis. *J Clin Epidemiol* 2015;**68**:860–9.
75. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandembroucke JP et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;**370**:1453–7.
76. Simunovic N, Sprague S, Bhandari M. Methodological issues in systematic reviews and meta-analyses of observational studies in orthopaedic research. *J Bone Joint Surg Am* 2009;**91**(Suppl_3):87–94.
77. Cameron C, Fireman B, Hutton B, Clifford T, Coyle D, Wells G et al. Network meta-analysis incorporating randomized controlled trials and non-randomized comparative cohort studies for assessing the safety and effectiveness of medical treatments: challenges and opportunities. *Syst Rev* 2015;**4**:147.
78. Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994;**140**:771–8.

79. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *Br Med J* 1998;**317**: 1185–90.
80. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AM. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess* 2000;**4**:1–154.
81. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;**342**:1878–86.
82. Shrier I, Boivin J-F, Steele RJ, Platt RW, Furlan A, Kakuma R et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007;**166**:1203–9.
83. Anglemeyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 2014;**4**:MR000034.
84. Egger M, Schneider M, Smith GD. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998;**316**:140–4.
85. Higgins JPT, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy* 2002;**7**:51–61.
86. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999;**354**:1896–900.
87. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;**4**:e78.
88. Moher D, Simera I, Schulz KF, Hoey J, Altman DG. Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research. *BMC Med* 2008;**6**:13.
89. Wen J, Ren Y, Wang L, Li Y, Liu Y, Zhou M et al. The reporting quality of meta-analyses improves: a random sampling study. *J Clin Epidemiol* 2008;**61**:770–5.
90. Gianola S, Gasparini M, Agostini M, Castellini G, Corbetta D, Gozzer P et al. Survey of the reporting characteristics of systematic reviews in rehabilitation. *Phys Ther* 2013;**93**:1456–66.
91. Page MJ, McKenzie JE, Kirkham J, Dwan K, Kramer S, Green S et al. Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. *Cochrane Database Syst Rev* 2014;**10**:MR000035.
92. Peters JPM, Hooft L, Grolman W, Stegeman I. Reporting quality of systematic reviews and meta-analyses of otorhinolaryngologic articles based on the PRISMA statement. *PLoS One* 2015;**10**:e0136540.
93. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med* 2016;**13**:e1002028.
94. Cullis PS, Gudlaugsdottir K, Andrews J. A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. *PLoS One* 2017;**12**:e0175213.
95. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Br Med J* 2009;**339**:332–336.
96. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;**4**:1–9.
97. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med* 2015;**162**:777–784.
98. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G et al. Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement. *JAMA* 2015;**313**:1657–1665.
99. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG et al. PRISMA harms checklist: improving harms reporting in systematic reviews. *Br Med J* 2016;**352**:i157.
100. Guise J-M, Butler ME, Chang C, Viswanathan M, Pigott T, Tugwell P et al. AHRQ series on complex intervention systematic reviews—paper 6: PRISMA-CI extension statement and checklist. *J Clin Epidemiol* 2017;**90**:43–50.
101. Guise J-M, Butler M, Chang C, Viswanathan M, Pigott T, Tugwell P et al. AHRQ series on complex intervention systematic reviews—paper 7: PRISMA-CI elaboration and explanation. *J Clin Epidemiol* 2017;**90**:51–58.
102. Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the Endorsement of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS One* 2013;**8**:e83138.
103. Tunis AS, McInnes MDF, Hanna R, Esmail K. Association of study quality with completeness of reporting: have completeness of reporting and quality of systematic reviews and meta-analyses in major radiology journals changed since publication of the PRISMA statement? *Radiology* 2013;**269**:413–26.
104. Riado Minguez D, Kowalski M, Vallve Odena M, Longin Pontzen D, Jelicic Kadic A, Jeric M et al. Methodological and reporting quality of systematic reviews published in the highest ranking journals in the field of pain. *Anesth Analg* 2017;**125**: 1348–54.
105. Pussegoda K, Turner L, Garrity C, Mayhew A, Skidmore B, Stevens A et al. Systematic review adherence to methodological or reporting quality. *Syst Rev* 2017;**6**:131.
106. Page MJ, Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement and extensions: a scoping review. *Syst Rev* 2017;**6**:263.
107. Zhang Z-W, Cheng J, Liu Z, Ma J-C, Li J-L, Wang J et al. Epidemiology, quality and reporting characteristics of meta-analyses of observational studies published in Chinese journals. *BMJ open* 2015;**5**:e008066.
108. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;**7**:1–7.
109. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;**62**:1013–20.
110. Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. *J Clin Epidemiol* 2015;**68**:574–83.
111. Public Policy Committee, International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practice (GPP). *Pharmacoepidemiol Drug Saf* 2016;**25**:2–10.
112. Alba S, Mergenthaler C. Lies, damned lies and epidemiology: why global health needs good epidemiological practice guidelines. *BMJ Glob Health* 2018;**3**: e001019.